Preparation, validation and application of combinatorial phagemid libraries displaying large peptides

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)
genehmigte
D i s s e r t a t i o n

von Erik Elmar Heinrich Pollmann
aus Düsseldorf

1. Referent: Prof. Dr. John Collins
2. Referent: Prof. Dr. Stefan Dübel
eingereicht am: 28.10.2009
mündliche Prüfung (Disputation) am: 12.02.2010

Druckjahr: 2010

Contents

Abbreviations

General abbreviations

Amp	Ampicillin
APS	Ammoniumpersulfate(-solution)
bp	Base pairs
cfu	Colony forming units per mL
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Mixture of desoxynucleotide triphosphates
	dATP, dTTP, dGTP and dCTP in equimolar concentrations
ddH_2O	Double deionized Water
dsDNA	Double stranded DNA
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
ELISA	Enzyme Linked Immuno Sorbent Assay
h	Hour
HRP	Horseradish peroxidase
min	Minute
mRNA	Messenger RNA
NaN_3	Sodium azide
NEB	New England Biolabs
OD_{600}	Optical density at a wavelength of 600 nm
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pfu	Plaque forming units per mL
rpm	Revolutions per minute

RT	Room temperature
SDS	Sodium dodecyl sulfate
ssDNA	Single stranded DNA
TCA	Trichloroacetic acid
TEMEDN	N,N',N'-Tetramethyl-ethylendiamin
T-PBS	Tween-20 (.. %) PBS buffer
Tris	Tris-(hydroxymethyl)-aminomethane
UV	Ultra violet

Amino acid codes

A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartic Acid	P	Pro	Proline
E	Glu	Glutamic Acid	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

1 Introduction

1.1 The challenge of molecular interactions

Molecular interactions between proteins and different ligands control most biological functions in organisms. These interactions are highly specific and thus permit an extremely intricate regulation of cellular functions. The identification of molecules involved in these specific interactions, as well as of molecules that might interfere with these complex cellular processes, forms the scientific basis of our understanding of differences between health and disease on the molecular and cellular level and its translation into pharmaceutical research.

Many different approaches have been developed to analyze the modes of interaction and to identify or develop novel ligands that can interact in a desired fashion.

One approach, called "rational design", uses computer-aided analysis and modeling of molecular structures to predict modes of interaction and alterations in the amino acid sequence of existing (known) molecules that might improve interactions (Shaikh et al., 2007). This method has been used in numerous experiments (e.g. Pisabarro and Serrano, 1996; Russell, 1998); however, it requires intimate knowledge of the three dimensional structures and the epitopes involved in the interaction in order to be able to predict the effect of changes. While this knowledge is available for an increasing number of macromolecules within bacterial and eukaryotic cells, the correlation of structure and function is still largely unknown. This limits the applicability of the "rational design" method.

A different approach, namely the empiric approach, sometimes ungenerously coined "irrational design", instead mimics the natural evolutionary process: In nature, all organisms are believed to have evolved from numerous generations of diversification (mutation) and selection of the "fittest" subjects, starting from a divergent "pool".

Chemical and biological methods are now applying this principle on the molecular level to evolve the function of molecules. Libraries that consist of up to 10^{10} different molecules are subjected to a selection process in which the "best" molecules are selectively enriched on the basis of an experimental criterium and finally identified as "best candidates" or "leads" for further development.

1.2 Biological libraries used for the identification of novel ligands by affinity selection

Although selection from libraries is performed with many different target properties in mind, e.g. for the directed evolution of enzymes to generate new variants, one of the most common applications is the selection for ligands that bind to a given target with a particularly high affinity. Identification of such novel high affinity ligands subsequently allows their use as highly sensitive biological probes for the detection of their respective target molecules and the analysis of interaction modes including competition for or blocking of receptor epitopes, thus being of high diagnostic and therapeutical interest.

Importantly, affinity selection allows the actual physical enrichment of ligands that have bound to a target of interest. Competition elution with compounds that bind to specific epitopes on the target molecule can therefore be used to characterize which epitopes are being bound, e.g. different ligands binding to different epitopes, or, as is often the case, can be used to select ligands binding to a specific desired epitope. This method has been successfully used e.g. to select for ligands binding to a substrate binding site or an allosteric activation site of an enzyme (see also Clackson and Wells, 1994).

1.2.1 Different presentation systems

During the screening process of biological libraries, the physical linkage of *genotype* — the DNA sequence encoding a protein — and *phenotype* — the gene product — is always required. As in Darwinian evolution, the potential benefits of an improved molecule would be lost without the blue print — i.e. the encoding gene.

All systems for *in vitro* evolution must as well link genotype and phenotype in some way, utilizing different methods of coupling the presented molecule(s) to the respective gene(s) encoding them. A number of systems which have been developed and employed for "display" screening/selection are described in the following sections.

1.2.1.1 Cell surface display

Cell-surface display is a method which uses living cells as a basis for the linkage of phenotype and genotype. With this method, proteins of interest are usually fused to a membrane protein so that they are presented on the cell surface (Lee et al., 2003; Olsen et al., 2000). Various cells, including *E. coli*, yeast and insect cells, are used for this purpose. This method has the advantage of being able to use the full cell arsenal of chaperones to ensure complete correct folding of the target protein, and in the case of eukaryotic cells also allows post-translational modifications (e.g. glycosylation).

Cell surface display makes the use of fluorescence-activated cell sorting (FACS) as a detection tool possible.

However, one has to deal with the complex system of a whole cell. This limits the conditions under which the system can be applied. The number of different clones is generally limited by the smaller size of the libraries.

Cell surface display is often used to present whole proteins and screen for protein *function*, although affinity selections with cell-surface based systems have also been used (e.g. by selecting for binding of fluorescence-marked antigens to achieve antibody maturation, see van den Beucken et al. (2003); Weaver-Feldhaus et al. (2004). Read Siegel (2009) for further examples).

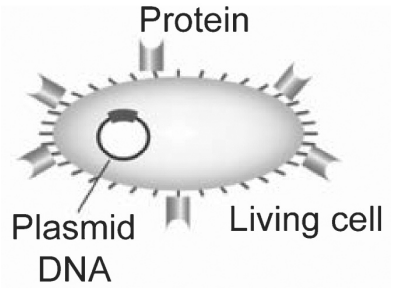


Figure 1: Cell surface display.
In cell-surface display, the proteins of interest are displayed on the surface of a living cell by anchoring via integrated membrane proteins. The gene encoding the fusion protein of interest and the anchor protein is encoded in a plasmid (from Matsuura and Yomo, 2006).

1.2.1.2 Ribosome display

In this method, a synthetic DNA library is expressed using an *in vitro* transcription+translation system. The occurring ribosomal complexes of mRNA, ribosome and nascent translated protein are stabilized to ensure a linkage of phenotype (the expressed peptide/protein) and the encoding gene (mRNA) (Hanes and Plückthun, 1997).

This stabilized state results in a non-covalently-bound complex which is relatively fragile and limits screening conditions to a narrow set of temperature and salt conditions. Modified protocols aim to enhance the stability, e.g. by expressing a fused MS2 coat-protein that will bind to a specific binding motif, which was originated from the bacteriophage MS2 (Sawata and Taira, 2003).

The main advantage of this method is that no DNA transformation step is required. High complexities of up to 10^{12} could be generated, (see Gersuk et al., 1997; He and Taussig, 2002; Mattheakis et al., 1994; Zhao et al., 2009).

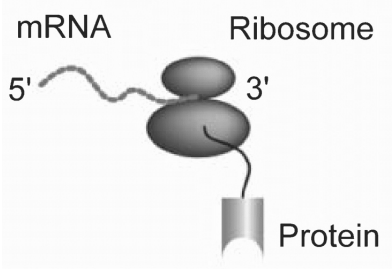


Figure 2: Ribosome display.
Ribosome display is based on the idea that ribosome complexes consisting of a nascent chain, ribosome and mRNA can be generated, establishing a link between the protein and the encoding gene (from Matsuura and Yomo, 2006).

1.2.1.3 mRNA display

In mRNA display, the mRNA and the encoded protein are linked covalently (in contrast to the non-covalent complex of ribosome display) via puromycin. The linking is achieved by having encoding mRNA that is bearing puromycin at its 3' end, which mimics aminoacyl tRNA. Puromycin will be covalently attached to the C-terminus of the protein once it enters the ribosome (Nemoto et al., 1997; Roberts and Szostak, 1997).

The resulting fusion molecule can be screened under any conditions that leave the RNA as well as the protein intact; reverse transcription of the RNA leads to a RNA/DNA hybrid which is more stable (Takahashi and Roberts, 2009).

The advantage of this technique is the large possible number of variants that can be obtained, since there is no limiting DNA transformation step; high complexities of up to 10^{12} could be generated (e.g. in the evolution of single-chain antibodies Fukuda et al., 2006).

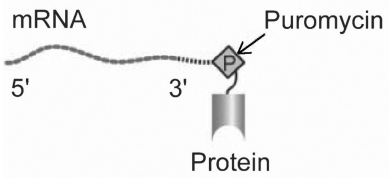


Figure 3: mRNA display.
mRNA display links the mRNA covalently to the encoded protein via puromycin, which is linked to the mRNA at the 3' end (from Matsuura and Yomo, 2006).

1.2.1.4 *In vitro*-compartmentalization

In vitro compartmentalization uses emulsions of water-in-oil: The aqueous core of each droplet holds a single gene encoding the protein of interest as well as an *in vitro* translation medium. The genotype-phenotype linkage occurs by encapsulating both the gene and the resulting encoded proteins (Bertschinger and Neri, 2004; Tawfik and Griffiths, 1998).

Griffiths and Tawfik (2003) achieve the linkage of gene and translated protein by embedding microbeads in the droplets, capturing DNA as well as protein via specific tags.

A method to *covalently* bind the encoding sequence to the resulting protein uses the M.hae III domain (Hae III DNA-methyltransferase of *Haemophilus aegypticus*) fused to the proteins of interest. The M. hae III domain can react with a modified methylation target DNA sequence to form a covalent bond (Chen et al., 1991). The DNA template encoding the M. hae III domain and the fusion protein carries this DNA modification, resulting in a DNA molecule "displaying" its covalently bound gene product.

After completed translation one can break up the compartments later and screen the proteins directly (Bertschinger et al., 2007; Griffiths and Tawfik, 2003).

This method currently is the only way to do an *in vitro* selection of enzyme activity.

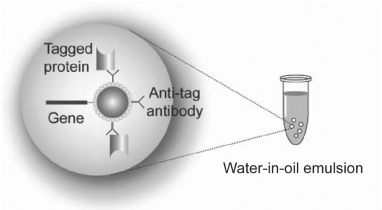


Figure 4: *In vitro*-compartmentalization.
In vitro compartmentalisation utilises the aqueous core of water-in-oil emulsions for *in vitro* translation. Proteins are expressed inside this core, and could also be immobilized on microbeads which the DNA could be fixed on. Expressed proteins are then immobilized on the same beads via an anti-peptide tag antibody, which is also immobilized on the beads (from Matsuura and Yomo, 2006).

1.2.1.5 Phage display

Phage display is one of the most commonly used methods for evolutionary experiments. Various types of phage are used, including the lambda phage, T7 phage, and filamentous phage. The most widely used system is that using the M13 filamentous phage (Smith, 1985). With this system, the protein of interest is fused to a coat protein of the phage and displayed on phage surface. The encoding gene is encapsulated inside the phage.

Phage display is typically used to select for binding interactions, such as from libraries of antibodies, antibody fragments, proteins using other scaffolds or peptides.

This system will be explained in more depth in section 1.3.

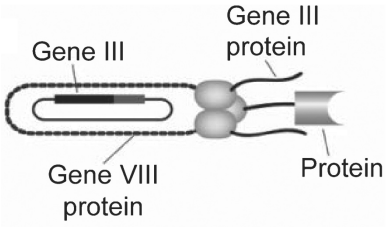


Figure 5: Phage display.
In this system, a protein of interest is displayed as a fusion protein with a phage coat protein on the surface of the phage, the gene encoding the protein is packaged inside the phage (from Matsuura and Yomo, 2006).

1.2.2 Structural considerations

In libraries that are designed to yield ligands with high affinities to specific varied targets, it has often been shown that ligands with more rigid structures show higher affinities. Be it the use of antibody Fab fragments (Knappik et al., 2000) or also multivalent antibody fragments (Plückthun and Pack, 1997), which possess a structure that nature has evolved as a scaffold, or the use of constrained proteins/peptides, a rigid structure often exhibits higher affinity than a more flexible chain of amino acids, due to the lower entropic price of binding caused by larger potential interaction surface.

Structural stability of the bound conformation also affects the specificity of binding, because of the inverse correlation between stability and the accessibility to alternative binding conformations (Ladner, 1995).

This is the reason why a number of different structural elements are used for stabilization. They are used to stabilize binding motifs as well as (and mainly) for the stabilization of a supporting scaffold, on which the variable region of the library is displayed.

Many different scaffolds have been used and are in use in libraries where the designers considered that reducing the entropy of the ligands will lead to higher affinities (other than libraries made from natural antibodies). Examples of such library scaffolds are

the following: The lipocalin derivatives that are used by the company *Pieris* (Skerra, 2000, 2001); libraries based on protease inhibitors (Roberts et al., 1992; Röttgen and Collins, 1995); or "affibodies" (based on the α-helical bacterial receptor domain Z from staphylococcal protein A) used by the company *Affibody* (Nord et al., 1997).

A limitation of these libraries often seems to be their specificity for a single target or a family of similar targets, since one would need to create a custom library for each target (e.g. anticalins that enclose small compounds as digoxigenin, but would not be suitable to interact with larger proteins). It must also be considered that, by prejudicing a starting structure, other, possibly more interesting ligands are removed from consideration.

In many cases where people want to create more general libraries, they do not go for pre-planned scaffolds, but just introduce some form of rigidity. Smaller domains such as zinc-finger domains (Bianchi et al., 1995), coiled-coil peptides or single helices (Houston et al., 1996) were used to present conformationally uniform peptide libraries.

While uncommon to appear in short peptides of <10 aa length, the formation of α-helices or β-sheets (for a review, see Binz and Plückthun, 2005) in a presented library would add a more defined shape that can interact with a target protein more specifically.

The smallest feature used often to (intentionally) introduce limited rigidity is the use of disulfide bridges between cysteines that constrain shorter or longer sequence motifs (Karlsson et al., 2004; McLafferty et al., 1993). Nature uses cysteine-rich mini proteins (for a look at the structure of knottins and short-chain scorpion toxins, see Combelles et al., 2008) to provide structure as well, cysteine knots as scaffolds have been used successfully (e.g. Christmann et al., 1999; Kimura et al., 2009), and the company *Amunix* (*California, USA*) uses short proteins with high cysteine content to provide rigidity for its display libraries.

1.3 Phage Display

The display of proteins and peptides on the surface of filamentous phage — *phage display* — is an *in vitro* selection method that allows the selection of polypeptides with desired properties from a large collection of variants. A gene encoding a peptide or protein of interest is fused to the gene of a phage coat protein, and after production and packaging of the phage particles, one gets phages that display the encoded protein *and* also contain its gene, providing the link between phenotype and genotype.

This technique allows phage libraries to be subjected to a selection step (e.g. affinity chromatography). Recovered clones can be identified by sequencing and re-grown for further selection rounds.

Since the initial description of this method by Smith (1985), phage display has become an established and well characterized method for the identification of polypeptides with

novel properties — examples are the isolation of peptides that bind to the target receptors for interleukin-1 (IL-1) (Yanofsky et al., 1996), thrombopoietin (Cwirla et al., 1997), or erythropoietin (EPO) (Wrighton et al., 1996) with affinities in the low micromolar to nanomolar range (see Clackson and Wells, 1994; Sidhu, 2000; Sidhu et al., 2000, for reviews).

The following sections detail the biology of filamentous phages, the different ways to present polypeptides via phage display, the considerations for selection processes utilizing phage display and different applications where phage display is actually used.

1.3.1 Biology of filamentous phage

1.3.1.1 The phage particle

Filamentous phages and infect Gram-negative bacteria.

Filamentous phages constitute a large family of bacterial viruses, usually contain a genome of single-stranded DNA, infecting a great variety of gram negative bacteria. The best-characterized filamentous phages are those that infect *E. coli*, which are the bacteriophages *f1*,*M13* and *fd*. These phages are collectively referred to as Ff phages. Their genomes are more than 98 % identical and their gene products interchangeable (Beck and Zink, 1981; van Wezenbeek et al., 1980). Unlike most other bacteriophages, filamentous phages are temperate phages, they are produced and secreted from infected bacteria continuously without lysing the cell (although a decrease in the rate of cell growth is seen in the infected cells).

The Ff phage genome consists of about 6400-nucleotide ssDNA; it holds 9 genes encoding 11 proteins and a major non coding "intergenic region", which contains the site of origin for the synthesis of the viral and complementary DNA strands as well as a site of initiation for phage particle assembly (packaging signal, *PS*) (Webster, 1996).

The genes can be divided into three groups according to their function: The first group encodes for DNA replication proteins (pII, pV, and pX), the second group encodes capsid proteins (pIII, pVI, pVII, pVIII, and pIX), while the third group encodes proteins that are required for the secretion of the phage proteins and the membrane associated assembly process of the phage particles (pI, pIV, and pXI).

The phage genome is encapsidated into a long cylinder of about 6.5 nm and a length of about 930 nm (depending on genome size, Specthrie et al. (1992)). The hollow tube consists of several thousand copies of the pVIII major coat protein (about 2700 copies per phage). On the "blunt" end of the phage, 3–5 copies each of pVII and pIX are located (with pVII being less exposed, Endemann and Model, 1995), while the other end holds about five copies of pIII and pVI (with pVI being less accessible).

The pIII protein (Figure 6) is vital for the infectivity of the phage, as well as for the

assembly of the phage particle. Its three domains (D1, D2 and D3) are connected by three glycine rich linkers. The domains D1 and D2 are important for the infectivity of the phage, while D3 is buried within the phage. D3 is necessary and sufficient for pIII to be incorporated into the phage and to mediate termination of phage assembly and release of the phage from the host cell (Rakonjac et al., 1999).

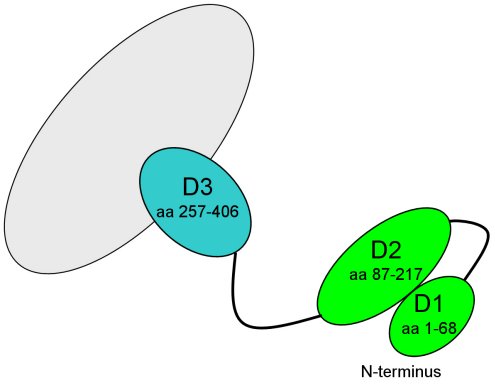


Figure 6: Domain structure of the pIII protein in filamentous phages.
The domains D1, D2 and D3 are linked by glycine rich linker regions. D1 is necessary for *E. coli* membrane penetration, while D2 is required for binding to the F-pilus of a host cell. D3 is buried within the phage particle (here grey) and also important for the assembly of the virion (see Figure 7)(Holliger and Riechmann, 1997).

The single-stranded DNA phage genome is oriented within the phage particle; its orientation is determined by the packaging signal site (*PS*) on the genome. The packaging signal, forming a stable hairpin, is positioned at the pVII-pIX end of the phage and is necessary and sufficient for the encapsidation of circular ssDNA into phage particles.

1.3.1.2 Infection of *E. coli* by Ff phages

Infection of *E. coli* by Ff phage normally begins when the D2 domain (see Figure 6) binds to the tip of the F conjugative pilus of an *E. coli* cell.

Since pili normally assemble and disassemble continuously, possibly also stimulated by phage binding, this brings the page close to the cell surface. Once the D2 domain binds the F pilus, the D1 domain is released from its normal interaction with D2, making it available to bind the host TolA protein. Three Tol proteins (Q, R and A), which are all integral cytoplasmic membrane proteins, are absolutely required for phage infection (Click and Webster, 1997). The exact mechanism of the phage penetration of the outer cell membrane is not yet characterized.

Once the pVIII capsid proteins have integrated into the inner membrane, the phage DNA (ssDNA, (+) strand, same polarity as mRNA) is released into the cytoplasm. Upon entry of the viral single-stranded phage DNA, host RNA and DNA polymerases

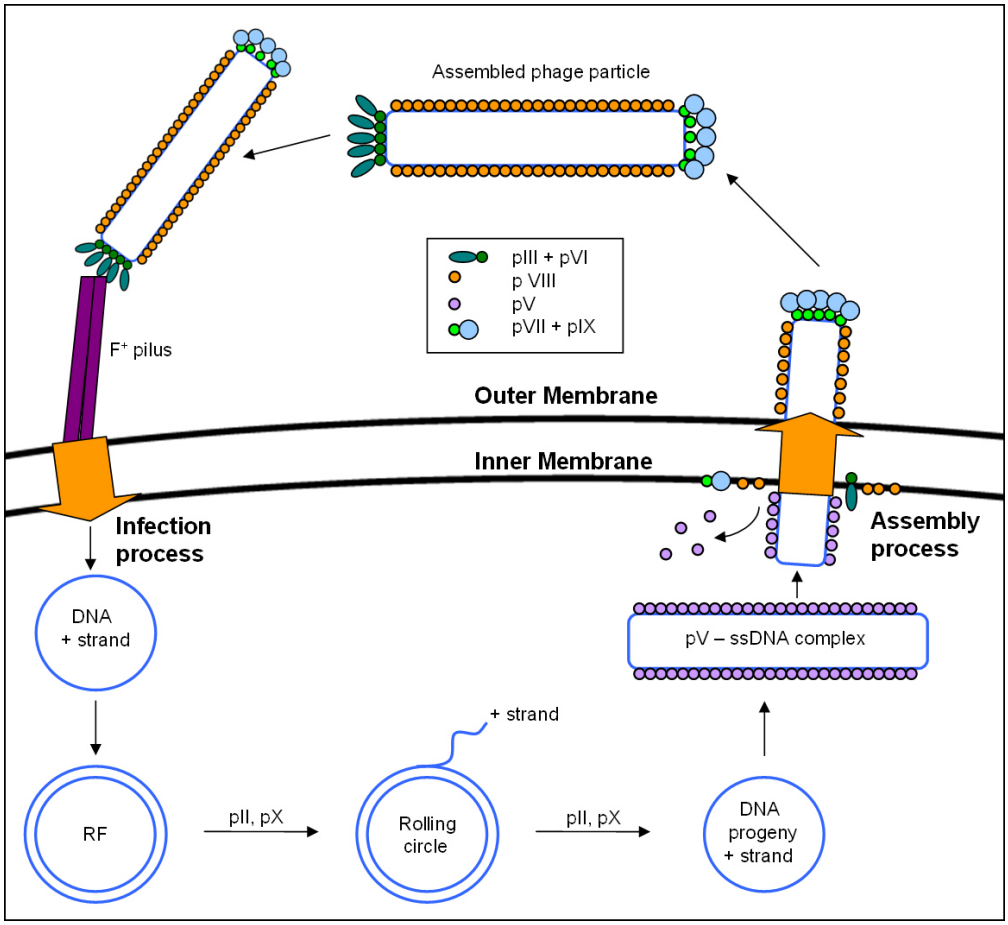


Figure 7: The life cycle of the Ff bacteriophage.
The phage enters the host cell through the F pilus, and the DNA (+ strand) is translocated into the cytoplasm. The bacterial enzymes synthesize the complementary strand, creating the replicative form (RF). The RF serves in a "rolling circle" reaction as template for the synthesis of new ssDNA (+ strand), while also phage proteins are transcribed and translated from the DNA. When pV concentration finally reaches a certain concentration, a complex of pV and the ssDNA is formed. A complex of pV and the DNA as well as pVII and pIX forms in the cell membrane, and in an translocation+elongation process through the membrane pV is replaced by membrane embedded pVIII. The assembly is terminated by the addition of the membrane-embedded complex of pIII-pVI, forming the end of the phage, upon which the complete phage is released from the cell.

and topoisomerases convert it into a double-stranded, super-coiled molecule that is called the "Replicative Form" (*RF*). The RF serves as template for phage gene expression. All phage proteins are expressed simultaneously, although several mechanisms ensure that each is produced at an appropriate rate.

The expression of pII particularly is necessary for further replication. pII, a site-specific nicking-closing enzyme, nicks the + strand of RF at a specific site in the non-coding region of the phage genome. The resulting 3' end is used by the host enzymes as a primer in order to synthesize a new (+) strand which is converted into a new RF DNA by bacterial enzymes.

The more RF DNA is produced, the more phage proteins are produced also, and proteins necessary for phage assembly get integrated into the cell membrane, while the proteins required for replication remain in the cytoplasm.

1.3.1.3 Phage assembly

The assembly of the phage particle takes place in the cytoplasmic membrane of the host cell; pI and pIX span the cytoplasmic membrane, while pIV forms a pore of about 8 nm ⌀ in the outer membrane through which assembled phage particles are secreted (Marciano et al., 2001).

The assembly is initiated by pV binding to the PS site of a newly synthesized circular (+) phage ssDNA. Together with pVII and pIX, which are most probably interacting with the PS, these proteins are assumed to create a complex in the membrane, forming the "tip" of a new phage.

The elongation of the phage involves successive replacement of pV dimers that cover the phage DNA by membrane-embedded pVIII and translocation of the DNA across the cell membrane. This process continues until the end of the phage DNA has been coated by pVIII.

The assembly is terminated by the addition of the membrane-embedded complex of pIII-pVI, forming the end of the phage, upon which the complete phage is released from the cell.

The phage production and assembly process is tolerated well by the cell, *E. coli* even continues to divide and grow after infection. The first phages appear already 10 min after infection of the host cell (at 37 °C), about 1000 phages are produced in the first hour.

Since cell growth continues, although significantly slower than uninfected cells, and with a lower maximal growth, on a bacterial lawn turbid plaques can be observed that contain infected cells and secreted phages (Model and Russel, 1988).

1.3.2 Principles of phage display

1.3.2.1 Phage proteins used for phage display

All five capsid proteins (pIII, pVI, pVII, pVIII and pIX) have already been used to display peptides or proteins, however, the by far most commonly used virion proteins for phage display are pVIII and pIII.

- **pVIII:** pVIII, the major capsid coat protein, is present in several thousand copies per phage particle (~2700 per wildtype phage). Displayed peptides are usually fused N-terminally, although one is usually limited to short peptides (6–8 amino acids), since larger sizes prevent packaging and would therefore not produce viable phages. The display of larger polypeptides requires the production of hybrid-pVIII phages by the use of a phagemid vector (see section 1.3.2.2).

- **pIII:** pIII is present in five copies at the "end" of the virion. It is used for most phage display fusions since it tolerates large insertions, and it can be used for monovalent presentation of a peptide/protein as well (see section 1.3.3.2). Fusions are usually N-terminal to pIII, between the leader sequence and the pIII protein.
  C-terminal fusions are rarely used, but would alleviate the requirement for in-frame expression between the leader sequence (required for secretion) and the N-terminus of pIII, which can help when cloning fragments of random length (e.g. cDNA).
  Although larger fusions are tolerated, they can have a dramatic impact on infectivity due to the importance of pIII for the infectivity of the phage (see section 1.3.1.2). This can also be alleviated by creating hybrid-pIII phages, where not all pIII proteins are modified (see section 1.3.2.2).

1.3.2.2 Phage display systems

Peptides/proteins can be displayed using vectors based on the natural Ff phage sequence — *phage vectors* — or using plasmid-based *phagemid* vectors that carry only the fusion protein gene and no other phage genes (Lowman, 1997).

Phage-display systems can be classified according to the arrangement of the coat protein genes. This is illustrated for fusions to pVIII and pIII in Figure 8, in which gene VIII is represented as a black block, gene III as a white block, the foreign DNA insert as a red block, and the foreign peptide as a red circle. In a type 3 vector, there is a single phage chromosome (genome) bearing a single gene III which accepts foreign DNA inserts and encodes a single type of pIII molecule. The foreign peptide encoded by the insert is theoretically displayed on all five pIII molecules on a virion. Type 8 is similar in regards to the pVIII coat protein, which is present at ~2700 copies per phage particle.

In a type 33 vector, the phage genome bears two genes III, encoding two different types of pIII molecule; one is recombinant (i.e., bears a foreign DNA insert) and the other wild-type. The resulting virion is a mosaic, its coat comprised of both wild-type and recombinant pIII molecules (the former usually predominating). This allows hybrid pIII proteins with quite large foreign peptides to be displayed on the virion surface, even though the hybrid protein by itself cannot support phage assembly. Type 88 has a similar expression pattern, although the size of fusion peptides is limiting (section 1.3.2.1).

A type 3+3 system differs from a type 33 system in that the two genes III are on separate genomes. The phagemid genome encodes only for the modified pIII fusion protein, while the so called "helper phage" provides the other phage proteins and also a wildtype pIII protein. Monovalent display is achieved by having the fusion pIII from the phagemid vector under control of a weak promotor, so that only a minority of the produced pIII in the bacterial host cell is the modified fusion protein, leading to a statistical monovalent presentation on the pIII protein of packaged phages.

The helper phage genome has a defective ori_{Ff}. This results in phagemid DNA being packaged preferentially into the assembled phage particles, to the detriment of the helper phage DNA (Russel et al., 2004); about 5–15 % of the packaged particles usually contain the helper phage genome, depending on the system used.

There have been also different systems developed that either modify or even replace the helper phage, see section 1.3.3.2.

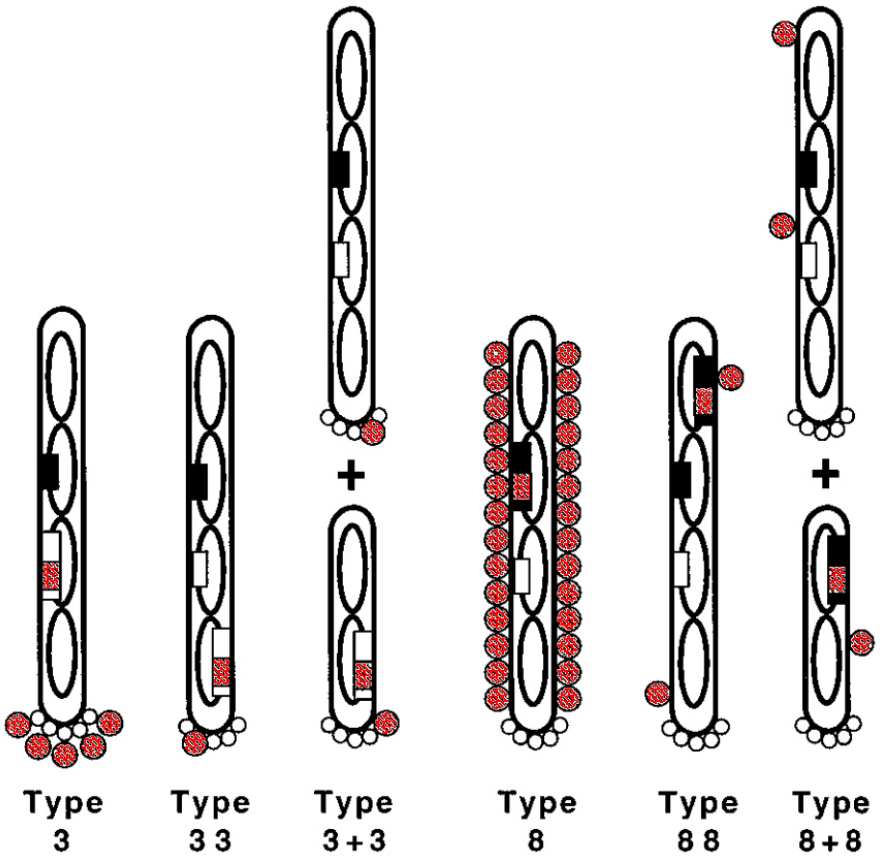


Figure 8: Types of phage display systems.
The long vertical ovals represent phage virions, and the shorter vertical ovals represent phagemid virions. The twisted line inside each virion represents the viral ssDNA, the segments encoding coat proteins pVIII and pIII being designated by black and white boxes, respectively. The red segments within these boxes represent foreign coding sequences spliced into a coat-protein gene, and the red circles on the surface of the virions represent the foreign peptides specified by these foreign coding sequences. The five white circles at one tip of the virions represent the N-terminal domains of the five pIII molecules. In type 8 systems, the foreign peptide is displayed on all copies of the major coat protein pVIII (2700 copies in wild-type virions), whereas in type 88 and 8+8 systems, only a minority of the pVIII copies display the foreign peptide (from Smith and Petrenko, 1997).

1.3.3 Screening of phage display libraries

Screening a phage display library consists of culling an initial population of phages presenting peptides or proteins to obtain a subpopulation with increased "fitness" according to a user-defined criterion , where affinity selection, screening the library for ligands that bind to a target of choice, is the most common application.

In most cases, the input to the first round of selection is a very large initial library. Typically 10^8-10^9 clones, each represented by 100–1000 particles on average, are applied in this initial challenge, resulting in phage titers of $> 10^8$. This gives a concentration for individual clones in the attomolar range, therefore binding in the first round will be close to a first order reaction, driven mainly by target concentration. The selected subpopulation is a relatively tiny fraction of the initial population (10^5-10^6 particles), "fitter" clones being over-represented due to favored enrichment.

This population can be "amplified" by infecting fresh bacterial host cells, so that each individual phage in the subpopulation is represented by millions of copies in the amplified stock. The amplified population can be subjected to further rounds of selection (perhaps accompanied by mutagenesis) to obtain an ever-fitter subset of the starting peptides. Already in the second round of panning, each clone is at least 1000-fold higher concentrated and more stringent washing conditions can be used.

1.3.3.1 The panning procedure

As mentioned, affinity selection for a target receptor is by far the most common selection pressure imposed on phage-displayed libraries. This selection is accomplished by minor modifications of standard affinity purification techniques that are commonly used in biochemistry. An advantage of the phage display system is the inherent robustness of phages that tolerate a wide range pH, are relatively temperature insensitive and resistant to many chemicals (like organic solvents, chaotropic agents as guanidinium chloride and many proteases), while still maintaining infectivity. This allows for a wide range of different selection environments.

A phage library is usually incubated with an immobilized target, which has been either covalently bound or just adsorbed to a treated surface. After the incubation, washing steps remove the non-binding phage. The number of washing steps and the stringency (e.g. salt concentration, detergents) determine how many phages remain bound to the target. After washing, the binding phages are eluted from the immobilized target and can be propagated. This cycle has to be repeated several times to enrich phage that bind strongly, until a significant enrichment of the eluted phage is observed. At this stage, individual clones are further analyzed by sequencing, and the clones are compared, with

the aim of distinguishing a consensus sequence.

Phages bind, are washed and retained phages are afterwards eluted. This method is also referred to as "sorting" or "(bio)panning" (Figure 9).

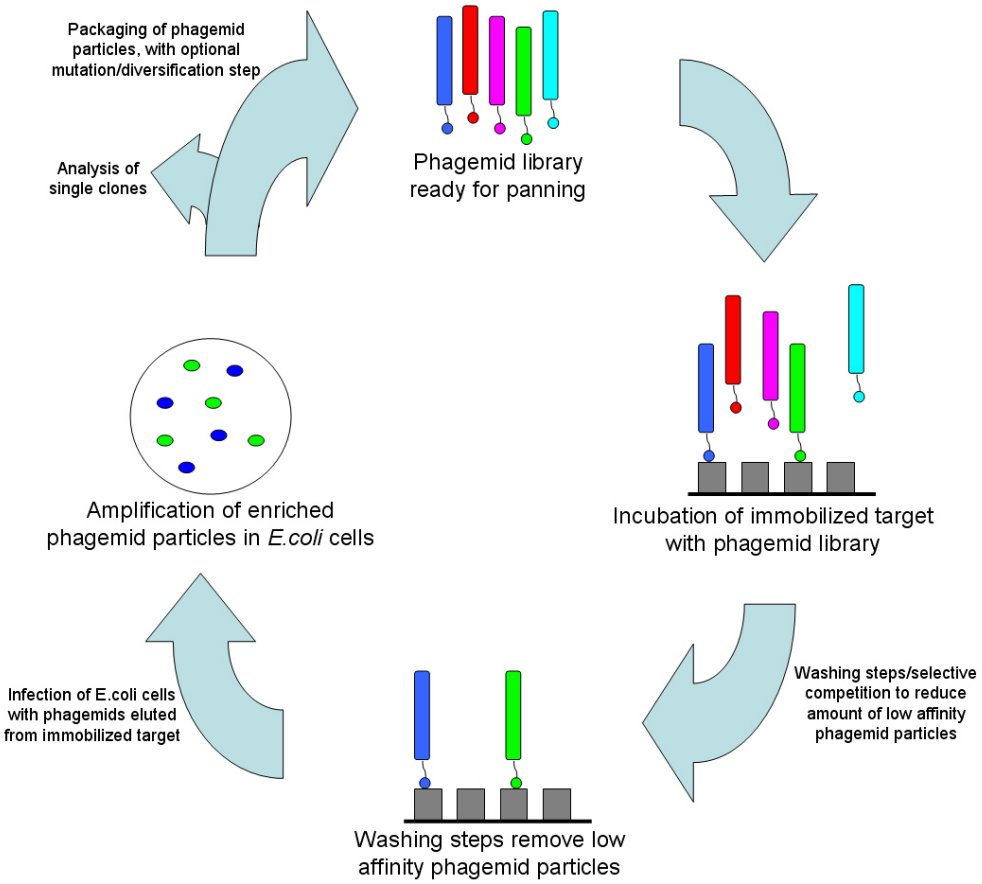


Figure 9: Affinity selection using phagemid libraries.

After the incubation of the target with the phage library, washing steps are performed in order to wash away preferentially particles showing low affinity to the target, leaving the strongly binding particles attached to the target.

For the elution of the retained phages, either specific or nonspecific conditions are used:

- **Nonspecific elution:** Nonspecific elution conditions are intended to weaken the target-peptide interactions without regard to their specificity. They exploit the high resistance of filamentous phages to denaturation by acidic buffers with pH down to 2.2 (Smith, 1985) or alkaline buffers such as 0.1 M triethylamine (up to pH 10.0 for 5–10 minutes) as reported by Harrison et al. (1996).

- **Specific elution:** Specific elution seeks to release phages that are bound to the target receptor's binding site, or in general a domain of interest, without releasing phages that are bound for some other reason, for example, by interaction with a contaminant, or with the immobilization surface. In competitive elution, a known soluble ligand for the receptor competes with phage for binding to the immobilized target molecule (e.g. Dybwad et al., 1995; Sioud et al., 1994). This method has the advantage of selecting for target and target-*site* specific phages.

  One can also use specific elution *during the challenge* to select for ligands that are binding to a *different* binding site (from personal communication with John Collins).

In the absence of knowledge about an appropriate method for elution, one can also just add *E. coli* host cells to the immobilized target-phage complexes directly after the washing steps (from personal communication with John Collins).

1.3.3.2 Considerations for screening of phage display libraries

Other than the general design of peptide libraries presented via phage display, as it is mentioned in section 1.4, there are also factors specific for phage display which must be kept in mind:

- **Library quality:** The quality of the encoded library is of crucial importance for the chances of isolating phages with desired properties. Developing a good strategy for the peptide design is important, but it must be validated later that the encoded peptides are still representative for that strategy. The steps of DNA modification, amplification and transformation can create a sequence bias that must be taken into account. Not only can bias arise through these steps, but also non-functional presentation, e.g. by insertion of stop codons in the DNA sequence of the fusion protein, leading to premature termination of translation.

- **Library complexity:** The transformation of DNA into cells before packaging into phage particles is a limiting step in phage display. Having a certain complexity in a library per design, *even discounting sequence/codon bias*, in order to cover the whole complexity with a 99 % probability, the number of clones in a library would have to be about 5× higher than the theoretical complexity of a library (Collins, 1997).

- **Valency of presentation:** The effect of either mono- or multivalent presentation (see section 1.3.2.2) on the enrichment of clones is also of great importance. While multivalent presentation usually yields more phage particles after a panning, it is

not as selective for high affinity ligands as a monovalent display system. This is due to avidity effects, which can make weaker binding effects appear stronger, even though the selection *ought* to be based on the affinity of a single presented molecule (Lowman et al., 1991).

In general, monovalent display is considered to enrich ligands with higher affinities; however, there are several protocols developed that consider a multivalent presentation in at least the first panning round as beneficial (due to the relatively low representation of each clone in the first round of panning, and thus the chance to lose promising candidates early on).

In order to be able to "switch" the mode of presentation, different approaches have been taken: By using an inducible promoter for the fusion protein used for presentation, leading to a higher fusion-protein to wildtype-protein ratio, by developing helper phages that do not encode a wildtype-protein themselves (e.g. Rondot et al., 2001), or even by eliminating the use of helper phage completely and using cells for phage packaging that carry different phage gene encoding plasmids, allowing control over the valency of presentation (Chasteen et al., 2006).

1.4 Peptide libraries as source for ligands

Most libraries used today for the isolation of high affinity ligands consists of recombinant/engineered proteins or antibodies/antibody fragments (like single-chain variable fragments, scFv); these libraries are screened for binding certain targets with high affinities and specificities and have become powerful tools for therapy, *in vivo* and *in vitro* diagnostics, drug target analysis and laboratory research (Binz and Plückthun, 2005).

However, proteins or antibodies are rather large molecules, a fact that can make production/expression of the ligands problematic. The cheaper process of complete chemical synthesis of interesting ligands usually is impossible or very difficult for larger molecules, which leaves researchers dependent on recombinant production, as there is no alternative path of validating results. Ligands that can be synthesized chemically are also desirable for potential pharmaceutic application so that no micro organisms would be required for production (Ladner et al., 2004).

Random peptide libraries have been used for many years to discover novel peptide reagents that have affinity for molecular targets. Phage libraries encoding six (Cwirla et al., 1990; Scott and Smith, 1990), seven (Scott et al., 1992), eight (Oldenburg et al., 1992) and fifteen (Devlin et al., 1990) random amino acids were used in screenings for high affinity ligands already more than 15 years ago.

Since then, libraries of small peptides have been mainly used for epitope mapping of antibodies, although the peptides themselves obtained from those experiments usually yielded only affinities at best in the low micromolar range, mostly due to binding motifs of only three or four amino acids length (noteworthy exceptions are high affinity ligands like HPQ for binding to streptavidin (Devlin et al., 1990) or the RGD motif for integrin binding (Ruoslahti, 1996)).

One reason to stop using of small peptide libraries was the potential complexity of a complete random library — a library of heptamers, using all 20 available amino acids, would already have a potential complexity of 20^7= ca. $1.28{\times}10^9$. A library of this size is already at the limit of what can be transformed into *E. coli*. Libraries with even longer peptides would not cover such a (potential) complexity completely.

Thus, since libraries of short peptides (< 10 aa) have already been used extensively, and libraries containing longer peptides sequences could not cover the potential complexity, researchers have not been using peptide libraries in searching for high affinity ligands in the last years.

However, there have also been ideas to use libraries containing longer peptides (e.g. Kay et al., 1993), due to the following ideas:

- Within long sequences, there are multiple short sequences because of a sliding window — e.g., within a 36-mer there are 31 different, adjacent hexamers.
- The peptides might be long enough to provide multiple contact sites necessary for a particular peptide-target interaction.
- The inserts might be large enough to assume a structure independent of the matrix/partner that is used to present the peptide — while short peptides are known to interact with the matrix quite often (e.g. with the phage surface in phage display systems).

Even though the complexities of libraries containing peptides of such length could not be covered in a library, these ideas have been applied to identify longer peptides binding specifically, as e.g. Dedman et al. (1993) with a library of 15mers, Sparks et al. (1994) with random 22mers, Adey and Kay (1996) with 26mers and Kay et al. (1993) with a library of random 38mers.

1.4.1 Increasing the diversity of a library

Having a large diversity is obviously a desirable property for a library. It increases the chances of there being a candidate with the desired properties in the library. This diversity can be achieved either by design in the original library (high sequence variety for libraries whose DNA template is synthesized; large variety of sources for library material obtained otherwise), or by increasing the diversity of a primary library later on (Gallop et al., 1994). The latter can be achieved by using methods that utilize mutation and, as is emphasized in the present investigations, recombination.

Error-prone PCR While the error-prone nature of the polymerase chain reaction (PCR) has been an issue almost since its original development, even the *Taq* polymerase with its relatively low fidelity is too accurate to be useful for the mutation of DNA libraries under standard reaction conditions.

Increases in error rates can be obtained by different ways — the most straightforward and popular method is the combination of introducing a small amount of Mn^{2+} (in place of the natural Mg^{2+} cofactor in PCR reactions) and using biased concentrations of dNTPs (Cirino et al., 2003).

A related method involves the synthesis of mixtures of small DNA molecules and is usually based on the incorporation of partially randomized synthetic small DNA cassettes into genes via PCR, directing the mutation to specific positions within the DNA template.

Mutator strains Low et al. (1996) developed a method that uses bacterial strains having a proof-reading-negative DNA polymerase III (*E. coli* mutator strain mutD5). Other strains have been developed that bear defects in one or several DNA repair pathways, leading to a higher mutation rate as well.

Genetic material passing through cells of these strains accumulates mutations at a vastly higher rate than usual. This is an effective and straightforward way of introducing mutations throughout a DNA construct.

The process of mutagenesis using mutator strains can be quite slow as the level of mutagenesis is controlled by the length of time the DNA spends in the strain (for a review of mutator strains, see Nguyen and Daugherty, 2003). In general, the rate of mutation is considered low, and since mutation is not limited to a library insert, a recloning of the library would be necessary to avoid mutations in the bacterium itself.

PCR based recombination of DNA sequences In contrast to the insertion of random mutations, natural evolution also exploits recombination to bring together advantageous mutations and separate out mutations that are less advantageous.

A number of different techniques of PCR based recombination utilizing partial homologies in DNA sequences is used:

The original technique of DNA shuffling (Stemmer, 1994a) is still one of the most popular tools in many optimization strategies. It allows to recombine a range of similar genes from various sources, or to combine selected point mutations in new combinations.

A number of other techniques have become available as well, each with own characteristics and uses, such as the staggered extension process (StEP), random chimeragenesis on transient templates (RACHITT) and different techniques based on iterative truncation for the creation of hybrid proteins (ITCHY) (all reviewed in Neylon, 2004).

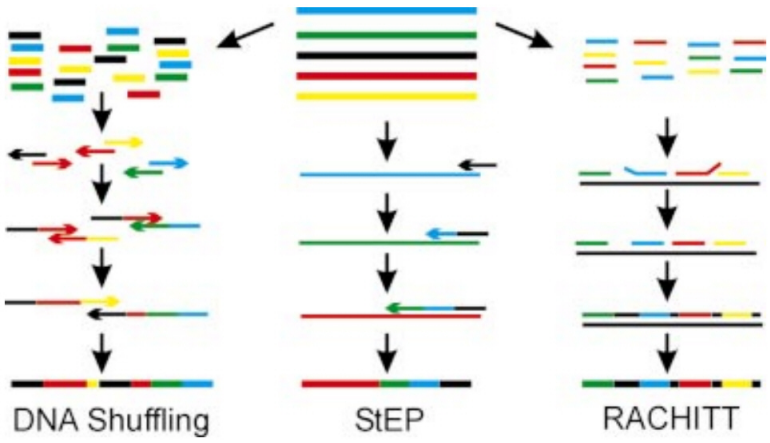


Figure 10: Homology based methods for recombining DNA sequences using PCR. DNA shuffling uses randomly cleaved parental DNA as at template in a self-priming PCR reaction; in StEP the DNA is not fragmented, but small segments are added to the end of a growing DNA strand in a series of very short extension steps. In RACHITT one parental DNA is used as a template where one strand in this template containing dUTP is generated. Fragments of the opposite strand of the other parental DNAs are then produced and annealed to the template. Non-annealed gaps are removed by exonuclease and filled using a DNA polymerase. These fragments are then ligated together and the template strand removed by endonuclease V digestion. The single strand is then converted to double-stranded DNA for further manipulations (from Neylon, 2004).

1.4.2 Cosmix-plexing

Cosmix-plexing was designed as a method to use recombination in DNA libraries without requiring large stretches of homologous regions. It allows the reassorting of hypervariable DNA regions, requiring only a minimum of two bases homology at a recombination site thus (Collins et al., 2001). This is compared to the previous PCR-based methods which require >70 % sequence homology on the DNA level to obtain efficient recombination.

Using this method, the potential diversity of the library is no longer limited to the complexity of the initial library. A Cosmix-plexing library contains one or more recombination sites for type IIs restriction enzymes between hypervariable DNA sequence *cassettes* (see Figure 11). These recombination sites allow the formation of novel com-

binations of cassettes to the left and to the right of the recombination site. Due to two defined base pairs without mirror symmetry at the recombination site, the cassettes will remain in the same order and orientation before and after the recombination. This is a highly efficient recombination process, in contrast to other methods that allow mutation or recombination. This is important for the resulting library quality in that no "junk" is produced from e.g. inverted sequences.

This technique can be used to increase the complexity of an initial library just by recombining the cassettes that already exist in the library (as shown in Figure 11), accessing new combinations of cassettes every time.

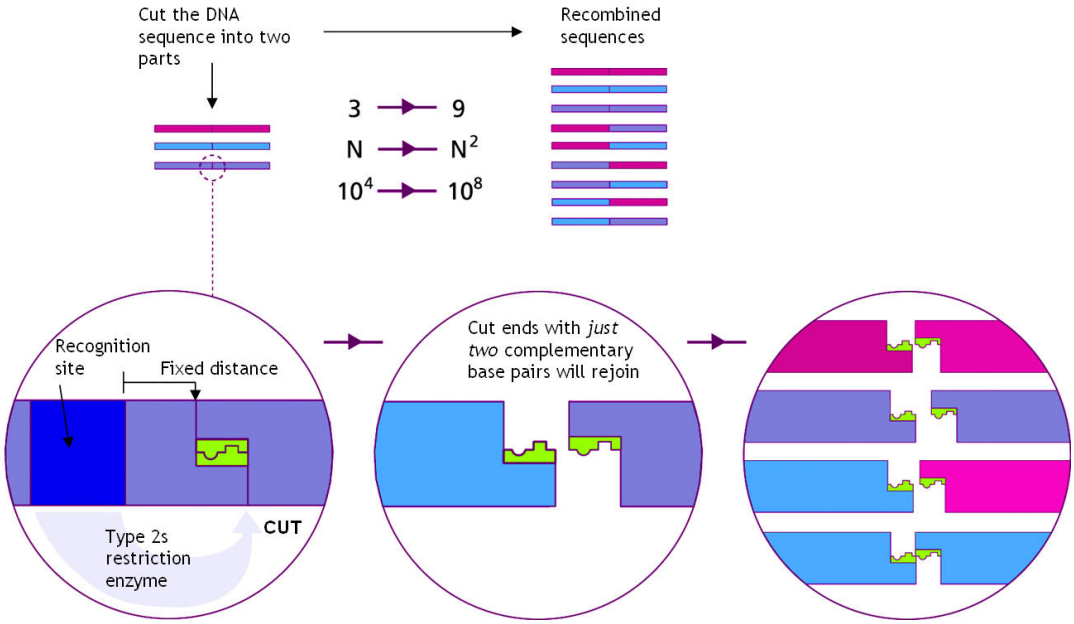


Figure 11: Principle of Cosmix Plexing. Peptide encoding DNA cassettes can be recombined using type IIs restriction enzymes, requiring only small fixed positions for ordered recombination instead of long homologous sequence stretches. Recombining a whole library this way gives access to new combinations every time (figure kindly provided by Prof. John Collins, Technical University Braunschweig, Germany).

One can also use Cosmix Plexing to recombine the DNA sequences of preselected clones from an early step in a screening (say, after 1–2 rounds of panning), thus recombining sequences that have already shown to be successfully enriched during the panning (Figure 12a). The method can also be used to "backcross" enriched clones with a library to increase the variety in defined cassettes, while keeping the enriched sequences of the remaining cassettes (Figure 12b). This way, it is possible to screen the library for sequences that are synergistic to the sequences in the kept cassettes. It can also be analyzed which cassette(s) of an enriched peptide are contributing most to enrichment by replacing either one or another sequence part. The Cosmix Plexing method will be explained in further detail in section 3.2.2.5.

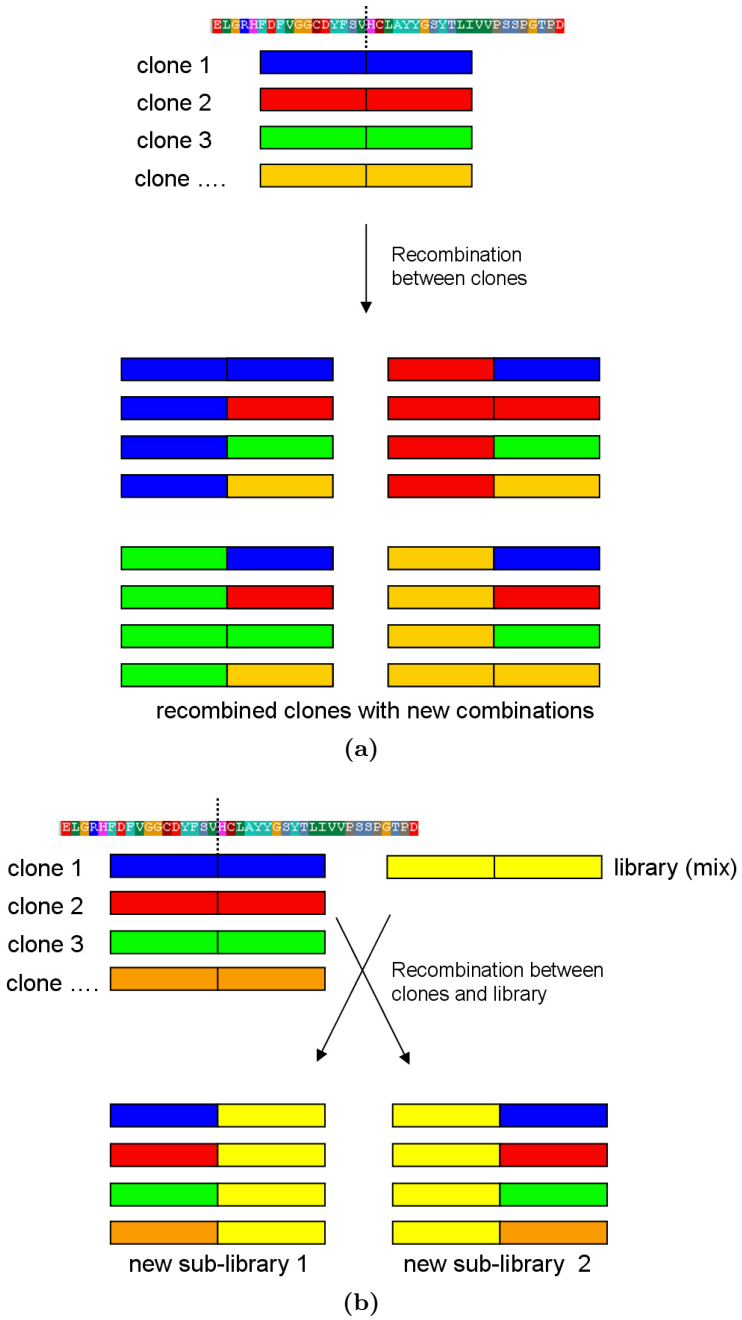


Figure 12: Further recombination possibilities offered by Cosmix Plexing:(a) Preselected sequences (blue, red, green, orange) can be recombined between each other to assemble new alternative variants. Recombination of sequences which have already been selected due to e.g. weak to medium affinities should offer a good probability of yielding sequences which possess even higher affinities to the selection target. (b) Preselected sequences (blue, red, green, orange) can be recombined with a whole library (yellow). Screening the obtained new library allows to evaluate the impact that the preselected cassette sequences have on affinity; it also allows to screen the recombination library (yellow) for motifs that have synergistic effects to the preselected motifs.

1.5 Aim of the work

Libraries utilizing phage display have been used for twenty years now; phage display libraries presenting peptides have been largely abandoned for other purposes than initial screenings. This is due to the fact that most will have an intrinsically low affinity to other small target ligands, but can be bind strongly to larger molecules having specific binding pockets (including antibodies). Their size however, in general limits potential interaction surface and the formation of secondary structures. Recent experiments show that library diversity increases probabilities of isolating high affinity ligands, which calls for structural diversity as well and is usually not obtained from short peptide libraries.

The advantages of peptide libraries include the easier chemical synthesis of interesting candidates and thus the possibility to generate peptides from D-amino acids (after a panning of L-peptides on a synthetic D-amino acid target molecule). The latter offers the chance of increased stability against proteolytic degradation and has already been used in the development of a gp41-targeting D-peptide (Eckert et al., 1999) and for identification of a cyclic D-peptide that interacts with cSRC SH3 domain (Schumacher et al., 1996).

The smaller size of peptides compared to antibodies makes direct application in cell penetration or generally as leads for research easier as well.

Mature protein domains that fold into stable secondary and tertiary structures have been found to consist of about 50–55 amino acids in average (Cunningham and Wells, 1997; Fung et al., 2008). There are also protein domains of smaller size, as the zinc finger domains (25–30 aa — Berg, 1988) or even the Trp-cage structure of 20 aa size (Neidigh et al., 2002, 2001). As a compromise between peptide length and the potential for secondary structures, the decision was made to use a polypeptide of 33 aa length (+ linker) in our library.

Recombination should allow for increase of diversity at *every* time of work with the library — at first in the beginning to increase the diversity of the *primary* library that is screened, or secondly later in the screening process where one could recombine clones that have been enriched due to their affinity to the target or just recombine high affinity clones with the primary library to vary/modify them.

The potential for recombination is vital to access the potential complexity of a peptide library of this length by backcrossing with either the primary library or a sublibrary of choice, and it offers also a tool for evolution or epitope screening of enriched candidates.

The recombination uses the technique of *Cosmix Plexing* (see section 1.4.2) to allow for ordered recombination. This reduces the amount of out-of-frame products which are a major disadvantage encountered using other recombination methods.

In this thesis the following steps are described which were used in building and evaluating the library:

- library design
- synthesis and cloning of the library, including composition analysis on DNA level and recombination functionality
- evaluation of peptide display
- evaluation of the library on different targets

The results are interpreted at each step. This is followed by a recapitulation of the whole work and major conclusions drawn with reference to the initial expectations and actual performance of this novel Cosmix Plexing technology in the *Conclusions* chapter. Finally projections are made as to how the project could be extended towards drug development.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

If not mentioned otherwise, all chemicals used in the context of this thesis were of analytical quality. Aqueous solutions were prepared with double deionized (Milli Q) water.

2.1.2 Software

Graphical and annotation work with plasmids was done using the program *pDRAW32 V1.1.97*. Analysis of sequencing data was done with *Chromas Lite V2.01*, alignment and sequence editing work was done with *BioEdit v7.0.9*.

This thesis was written in and typeset with LaTeX 2_ε, from the MikTex 2.7 distribution (www.miktex.org), using the editor WinEdt.

2.1.3 Bacteria, bacteriophages and plasmids

2.1.3.1 Bacteria

The following strains were used in the work for this thesis:

- *Escherichia coli* Top10F' (*Invitrogen*) (Top10F')
  genotype: F´{*lac*Iq, *Tn*10(TetR)} *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*Z ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G
- Top10F'λ (λ lysogen Top10F' strain) kindly provided by Prof. John Collins, Technical University Braunschweig, Germany

2.1.3.2 Bacteriophages

- Helper-phage M13K07 (*New England Biolabs*)

2.1.3.3 Plasmids

- Phagemid vector pROCOS 4/7, constructed by Peter Röttgen, kindly provided by *Cosmix GmbH, Braunschweig, Germany*

2.1.4 Media and antibiotics for cultivation of bacteria

The following media and antibiotics were used for the cultivation of bacteria, :

2×YT medium (*2YT*)	
Tryptone	16.0 g
Yeast extract	5.0 g
NaCl	5.0 g
Agar (for solid medium only)	20.0 g
ddH_2O	fill up to 1000 mL
adjust pH to 7.0 if necessary, then autoclave	

M9 salt solution (10×)	
Na_2HPO_4*2 H_2O	74.1 g
KH_2PO_4	30 g
NaCl	5.0 g
NH_4Cl	10.0 g
Agar (for solid medium only)	20.0 g
ddH_2O	fill up to 1000 mL
autoclave	

M9 minimal medium	
M9 salt solution (10×)	100 mL
Glucose (40 % w/v)	12.5 mL
CaCl (1 M)	100 µl
$MgSO_4$ (1 M)	1 mL
$FeCl_3$ (1 mM)	0.5 mL
Thiamin (10 mg/mL)	100 µl
ddH_2O	fill up to 1000 mL
The used solutions were autoclaved separately or filtered (0.2 µm) for sterilization (Thiamin).	

Antibiotics	
Ampicillin stock solution (50 mg/mL)	
Ampicillin (sodium salt)	2.5 g
Ethanol (70 % in ddH_2O)	50 mL
The final concentration in either liquid or solid medium to select for ampicillin resistance was **100 µg/mL**.	
Tetracycline stock solution (12.5 mg/mL)	
Tetracycline	0.625 g
Ethanol (70 % in ddH_2O)	50 mL
The final concentration to select for tetracycline resistance in liquid medium was **10 µg/mL** and **12.5 µg/mL** in solid medium.	
Kanamycin stock solution (50 mg/mL)	
Kanamycin	2.5 g
ddH_2O	50 mL
The solution was filtered (0.2 µm) for sterilization The final concentration in either liquid or solid medium to select for kanamycin resistance was **30 µg/mL**.	

2.1.5 Oligonucleotides

2.1.5.1 Oligonucleotides for library generation (section 3.2.2.2)

CPL19YS-2 BAS (132 bp):
5'- TTA CAT AGA GCT CGG ACG GCN TNH TNH TNH TNT SNV TNV TNV TNV TTM CNH TNH TNY TCA TTG CNH TNH TNH THM YNV TNV TNV TNC CNH TNH TNH TNH TNH CTC CAG CCC AGG TAC CCC GGA TAT CAT GGC -3'

JCOSTART3B-3 (18 bp):
5'-GCC ATG ATA TCC GGG GTA -3'

2.1.5.2 Primers used for mutagenesis PCR (section 3.2.1.1)

For the removal of the *Bce*AI recognition site at pos. 489 of pROCOS 4/7:
BCE-FX-1 (forward, 33 bp):
5'- CCC TCT CGA TGG CAC TTA TCC GCC TGG TAC TGA -3'
BCE-X-1 (reverse, 38 bp):
5'- GGC GGA TAA GTG CCA TCG AGA GGG TTG ATA TAA GTA TA -3'

For the removal of the *Bce*AI recognition site at pos. 1906 of pROCOS 4/7:
BCE-FX-2 (forward, 22 bp):
5'- GTG CTT TAC GAC ACC TCG ACC C -3'
BCE-X-2 (reverse, 26 bp):
5'- CGA GGT GTC GTA AAG CAC TAA ATC GG -3'

For the removal of the *Bce*AI recognition site at pos. 2758 of pROCOS 4/7:
BCE-FX-3 (forward, 28 bp):
5'- GGC CTA ACT ACG ACT ACA CTA GAA GGA C -3'
BCE-X-3 (reverse, 27 bp):
5'- GTG TAG TCG TAG TTA GGC CAC CAC TTC -3'

For the removal of the *Bce*AI recognition site at pos. 4075 of pROCOS 4/7:
BCE-FX-4 (forward, 30 bp):
5'- GGC AGC AGT GAC GTC GAT AGT ATG CAA TTG -3'
BCE-X-4 (reverse, 30 bp):
5'- CTA TCG ACG TCA CTG CTG CCA GAT AAC ACC -3'

For the removal of the *Bsr*DI recognition site at pos. 3210 of pROCOS 4/7:
ABSRD-1FX (forward, 22 bp):
5'- GGC CCC AGT GCT GCG ATG ATA C -3'

ABSRD-1 (reverse, 22 bp):
5'- CTC GCG GTA TCA TCG CAG CAC T -3'

For the removal of the *Bsr*DI recognition site at pos. 3384 of pROCOS 4/7:
ABSRD-2F (forward, 29 bp):
5'- GTT GCC ATA GCT GCA GGC ATC GTG GTG TC -3'
ABSRD-2Z (reverse, 23 bp):
5'- TGC CTG CAG CTA TGG CAA CAA CG -3'

For the removal of the *Bsr*DI recognition site at pos. 4012 of pROCOS 4/7:
ABSRD-3FX (forward, 28 bp):
5'- TTT ATC TTG TGC AAC GTA ACA TCA GAG A -3'
ABSRD-3X (reverse, 27 bp):
5'- CTG ATG TTA CGT TGC ACA AGA TAA AAA -3'

For the removal of the *Bsr*DI recognition site at pos. 4109 of pROCOS 4/7:
BSRDCURE-F (forward, 40 bp):
5'- GTC GAT AGT ATG CAA TTG ATA ATT ATT ATC ATT TGC GGG T -3'
BSRDCURE-R2 (reverse, 38 bp):
5'- CAA ATG ATA ATA ATT ATC AAT TGC ATA CTA TCG ACG TC -3'

2.1.5.3 Primers used for sequencing of the library insert in pEPO 8

JCOFSQ (binding at pos. of 4634 of pEPO 8,forward, 20 bp):
5'- TTC TAC AAC TTG CTT GGA TT -3'

JCORSQ (binding at pos. of 237 of pEPO 8,reverse, 20 bp):
5'- TCC AGA CGT TAG TAA ATG AA -3'

2.1.5.4 Primers used for detection of λ prophage integration into the *E. coli* genome (section 2.2.2.4)

Primer Lambda1 (forward, 22 bp):
5'- GAG GTA GCA GCG CGG TTT GAT C -3'
Primer Lambda2 (reverse, 18 bp):
5'- ACT CGT CGC GAA CCG CMC- -3'

2.1.5.5 Various primers used for plasmid sequencing and PCR synthesis of DNA fragments for PCR mutagenesis

ABPUE3-1 (binding at pos. 540 of pROCOS 4/7 - forward, 24 bp):
5'- AAT CCT TCT CTT GAA GAG TCT CAG -3'

ABPUE3-2 (binding at pos. 631 of pROCOS 4/7 - forward, 26 bp):
5'- TAC GGG CAC TGT TAC GCA AGG CAC TG -3'

ABPUE3-3 (binding at pos. 1145 of pROCOS 4/7 - forward, 24 bp):
5'- TTC CCA AAT GGC CCA AGT CGG TGA -3'

ABPUE4 (binding at pos. 2375 of pROCOS 4/7 - forward, 27 bp):
5'- ACA AAA ATC GAC GCV CAA GTC AGA GGT -3'

ABPUE5 (binding at pos. 2640 of pROCOS 4/7 - forward, 25 bp):
5'- ACT ATC GTC TTA AGT CCA ACC CGG T -3'

ABPUE7 (binding at pos. 3763 of pROCOS 4/7 - reverse, 23 bp):
5'- GCG GTA AGA TCC TCG AGA GTT TT -3'

BPUE-66 (binding at pos. 2885 of pROCOS 4/7 - reverse, 36 bp):
5'- GAT CAA AGG ATC TTC TTG GGA TCC TTT TTT TCT GCG -3'

XMN1171FOR (binding at pos. 1190 of pROCOS 4/7 - forward, 32 bp):
5'- CCT TTA ATG AAT AAT TTC CGT CAA TAT TTA CC -3'

NEW1170RVP (binding at pos. 1159 of pROCOS 4/7 - reverse, 24 bp):
5'- TTG ACG GAA ATT ATT CAT TAA AGG -3'

SEQ3606 (binding at pos. 3628 of pROCOS 4/7 - forward, 23 bp):
5'- GAT GCT TTT CTG TGA CTG GTG AG -3'

RSEQ4306 (binding at pos. 4285 of pROCOS 4/7 - reverse, 22 bp):
5'- GGC CAA AAA GCT CGC TTT CAG C -3'

F133 (binding at pos. 151 of pROCOS 4/7 - forward, 19 bp):
5'- CCC GGG TAC CCC GAT ATC A -3'

R1084 (binding at pos. 1064 of pROCOS 4/7 - reverse, 21 bp):
5'- TAG CAA GGC CGG AAA CGT CAC -3'

F1335 (binding at pos. 1358 of pROCOS 4/7 - forward, 24 bp):
5'- TTC TAC GTT TGC TAA CAT ACT GCG -3'

R1986 (binding at pos. 1967 of pROCOS 4/7 - reverse, 20 bp):
5'- GGA CTC CAA CGT CAA AGG GC -3'

F2548 (binding at pos. 2565 of pROCOS 4/7 - forward, 18 bp):
5'- CGT TCG CTC CAA GCT GGG -3'

R3398 (binding at pos. 3380 of pROCOS 4/7 - reverse, 19 bp):
5'- TGC CTG CAG CAA TGG CAA C -3'

F4027 (binding at pos. 4046 of pROCOS 4/7 - forward, 20 bp):
5'- GAG ACA CAA CAG ATC TGG CC -3'

R4648 (binding at pos. 4626 of pROCOS 4/7 - reverse, 23 bp):
5'- TCT AGA CCT CCT TGG TCG ACT TC -3'

2.1.6 Common buffers and solutions

Phosphate buffered saline (*PBS*), 10× stock	
KCl	2 g
KH_2PO_4	2.4 g
$Na_2HPO_4 \times 2H_2O$	14.4 g
NaCl	80 g
ddH_2O	fill up to 1000 mL
adjust pH to 7.4 if necessary (using either HCl or NaOH), then autoclave. Dilute with sterile distilled water to 1× to use.	

PEG/NaCl (16,7 %/3 M)	
PEG 8000	167 g
NaCl	0.2 g
ddH_2O	fill up to 1000 mL

2.2 Methods

2.2.1 Work with bacteria

2.2.1.1 Cultivation of bacteria

For liquid culture, *E. coli* cells were cultivated in baffled Erlenmeyer flasks filled up to 1/4th of the nominal volume with 2YT medium. When required, respective antibiotics were added. Incubation was done at 37 °C under shaking (160 rpm).

For cultivation on agar plates *E. coli* cells were grown on 2YT agar, containing appropriate antibiotics. Incubation was done at 37 °C.

2.2.1.2 Preparation of electrocompetent cells

The preparation of electrocompetent *E. coli* cells is following the protocol of Dower et al. (1988).

Bacteria from the strain to be made competent were streaked out on M9 minimal agar plates with appropriate antibiotics. A single colony was picked as inoculatum for an overnight culture in 2YT medium with selective antibiotics.

With 10 mL from this overnight culture, 1 L of prewarmed (37 °C) 2YT medium (with strain specific antibiotics) was inoculated and grown at 37 °C, 160 rpm, until an OD_{600} of 0,4–0,6 was reached. The flasks were finally cooled on crushed ice directly for 15 min.

The cell suspension was centrifuged using a pre-cooled GS3 rotor at 4,000 × g for 15 min. After careful decanting of the supernatant, the cells were resuspended in 1 L of precooled ddH_2O (0 °C), and recentrifuged using GS3 rotor centrifuge tubes at 4,000 × g for 15 min. The supernatant was decanted again and the cell pellet suspended in 500 mL of precooled ddH_2O (0 °C). This was followed by another centrifugation step (GS3 rotor, 4,000 × g, 15 min). The cell pellet remaining after decanting the supernatant was taken up in 20 mL of pre-cooled 10 % glycerol (0 °C) and then spun down in SS34 centrifuge tubes (precooled SS34 rotor, 4,000 × g, 15 min). After carefully discarding the supernatant, the pellet was suspended in 2 mL of pre-cooled 10 % glycerol, aliquoted in 50 µL aliquots and then deepfrozen in liquid nitrogen, followed by storage at -70 °C.

2.2.1.3 Storage of bacteria

For short term storage of up to two weeks, bacteria were stored on agar plates sealed with parafilm at 4–8 °C.

For longer term storage, 1 mL of bacterial cells (either from a fresh overnight culture or a cultivation up to a OD_{600} of 0.5–1) were mixed with 0.5 mL of sterile 87 % glycerol,

deep frozen in liquid nitrogen and then stored at -70 °C. Under those conditions, strains could be stored almost indefinitely.

2.2.2 Work with DNA

2.2.2.1 Isolation of plasmid DNA

For the isolation of plasmid DNA, preparational kits from *Qiagen* were used, either *QIAprep Spin Miniprep Kit* for small-scale preparations (1–5 µg plasmid DNA from 1–2 mL overnight cultures in 2YT medium), or *QIAGEN Plasmid Maxi Kit* for larger-scale preparations (>50 µg plasmid DNA from 50–200 mL overnight cultures in 2YT medium). The plasmid DNA isolation was performed according to the manufacturer's instructions.

2.2.2.2 Determination of DNA concentration

DNA concentration was determined by measuring the absorption of monochromatic light of a wavelength of 260 nm (A_{260}) in Hellma quartz glass cuvettes.

Background absorption was measured in a quartz cuvette filled with either ddH_2O or an appropriate buffer. DNA was then added to the solution used as background and the absorption was measured. The DNA concentration was calculated by multiplying the absorption with the dilution factor and the factor $\mathbf{50\,ng/\mu L}$.

$$\text{DNA concentration}[ng/\mu L] = A_{260} \times \text{dilution factor} \times 50\,ng/\mu L.$$

2.2.2.3 Polymerase Chain Reaction (PCR)

The PCR method was used to amplify DNA fragments and insert mutations by using selected oligonucleotides as primers. For the reactions, the *Phusion PCR kit* (*Finnzymes, Finland*) was used, using a high-fidelity DNA polymerase. The thermocycler used was a Primus Thermocycler Evolution 96 (*Clemens GmbH, Germany*).

If not mentioned otherwise, the following PCR reaction mix and protocol were used. The amount of DNA template used for the reaction was always «100 ng of DNA. For a negative control, 1 µL ddH_2O was added instead of the template DNA.

2.2.2.4 Detection of phage λ prophage integration into the *E. coli* genome

The Top 10F'λ strain was provided with the phage *Lambda* already integrated into the bacterial genome (lysogenic state). This integration was only assumed, though, so an insertion into the genome had to be proven before further work (the test was also used later on to check and differentiate between Top 10F' and Top 10F'λ strains).

PCR reaction mix	
Phusion HF 5×buffer	10 µL
Forward primer (10 µM)	2 µL
Reverse primer (10 µM)	2 µL
dNTP solution (2 mM each)	2 µl
DNA template	1 µl
Phusion polymerase	0.5 µl
ddH_2O	32.5 µl
Final volume	**50 µl**

PCR program		
Step	**Temperature**	**Duration**
pre-denaturation	98 °C	2:00 min
denaturation	98 °C	1:00 min
annealing	primer dependent	0:40 min
elongation	72 °C	1:00 min
postelongation	72 °C	3:00 min
The steps "denaturation" to "elongation" were repeated as cycles, dependent on the reaction yield, between 10 and 25 times.		

In order to integrate into the bacterial genome to reach the lysogenic state, phage Lambda encodes an integrase enzyme that mediates recombination between a short sequence of phage DNA (see Groth and Calos, 2004), the phage attachment site, *attP*, and a short sequence of bacterial DNA, the bacterial attachment site, *attB*. An integrated phage is flanked by two hybrid sites, *attL* and *attR*, each consisting of half *attP* sequence and half *attB* sequence (Figure 13).

The integration of lambda DNA can be detected by using PCR primers that are specific to the DNA sequence left of the *E. coli* attB site (Lambda1) and to the *Lambda* phage *int* gene right to the attB site after integration (Lambda2); this is illustrated in Figure 14.

A PCR reaction utilizing those primers will only give a PCR amplicon when the prophage is integrated. The expected product size (calculated from the two genome sequences, *E. coli* and phage *Lambda*) is 501 bp, so presence of this band after gel electrophoresis of a PCR reaction using genomic *E. coli* DNA as template could be seen as confirmation of *Lambda* prophage integration, absence of the band meant no *Lambda* prophage integration.

The PCR reaction was performed at an annealing temperature of 58 °C for 15 cycles. As DNA template, some bacterial cell material was used, diluted in water and boiled at 100 °C for 5 minutes. From this solution, 1 µL was used for a 50 µL PCR reaction.

Primer Lambda1 (forward, 22 bp):
5'- GAG GTA GCA GCG CGG TTT GAT C -3'
Primer Lambda2 (reverse, 18 bp):
5'- ACT CGT CGC GAA CCG CMC- -3'

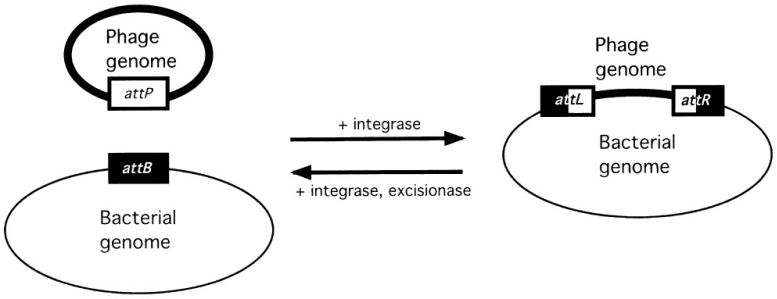


Figure 13: Phage genome integration/excision. The bold circle represents the phage genome, with the phage attachment site *attP*. The larger circle represents the bacterial genome, with the bacterial attachment site *attB*. In order to enter the lysogenic state, the integration reaction proceeds, resulting in the phage genome integrated into the bacterial chromosome, flanked by two hybrid *att* sites, *attL* and *attR*. Once the phage becomes lytic again, the reverse reaction occurs. The phage genome is excised from the bacterial chromosome, resulting in a phage genome with *attP* and a bacterial genome with *attB*(from Groth and Calos, 2004).

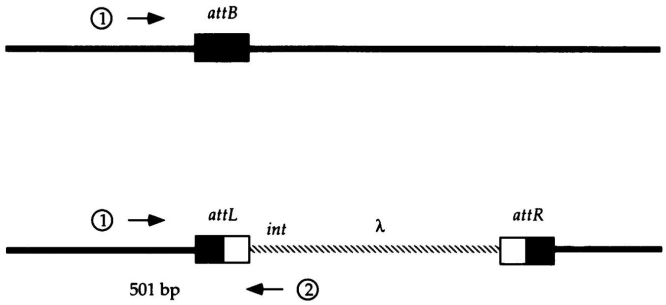


Figure 14: Detection of lambda prophage integration by PCR. The diagram shows the different att sites occurring in the genome of a λ nonlysogen *E. coli* (top) and λ lysogenic *E. coli* (bottom). Lines depict bacterial DNA (solid) and lambda DNA (hatched), with *att* sites represented by boxes. The respective locations and orientations for hybridization of PCR primers numbered 1 (primer Lambda1) and 2 (primer Lambda2) are indicated by arrow heads above and below the DNA lines. The fragment size of 501 bp produced by productive priming is noted in case of λ prophage integrated into the bacterial genome (from Powell et al., 1994).

2.2.2.5 DNA agarose gel electrophoresis

For the DNA separation in agarose gels, gel chambers from *Biometra* (Germany) were used — the Horizon 58 model for analytical, the Horizon 11.14 model for preparative gels. As running buffers, TAE and TBE were used. TAE was used for analytical and preparative gel runs, while TBE was only used for the high resolution separation of very small DNA fragments (below 150 bp).

The agarose gels (concentrations 0.8 %–3 % agarose w/v) were prepared by dissolving standard agarose (*Agarose GTQ, Carl Roth, Germany*) in the appropriate buffer by heating. After cooling to about 50 °C the gel was cast in the gel running chamber that contained pocket forming combs.

The DNA loading buffer used was *6X DNA Loading Dye* from *Fermentas, Canada*, DNA markers used were *GeneRuler DNA ladders* of different fragment sizes obtained from *Fermentas, Canada*.

DNA agarose gels were run at voltages between 60 V and 150 V, depending on gel chamber system and experimental factors. The power supply used was a PowerPack P25 (*Biometra, Germany*).

Staining of DNA in agarose gels after gel run was performed by mildly shaking the gel in an ethidium-bromide solution (10 mg/L) for 10 min, then washing out background in ddH_2O for 5 min.

The DNA was then detected using the *Intas Gel Jet Imager* gel documentation system (*Intas, Germany*), based on UV illumination and a filter for a wavelength of 312 nm, a digital video camera and computer image capturing.

TAE buffer 50×	
Tris base	242 g
Acetic acid	57.1 mL
0.5M EDTA	100 mL
ddH_2O	fill up to 1000 mL
dilute with ddH_2O to get 1×	

TBE buffer 50×	
Tris base	54 g
Boric acid	27.5 g
0.5M EDTA, pH 8.0	20 mL
ddH_2O	fill up to 1000 mL
dilute with ddH_2O to get 1×	

2.2.2.6 DNA polyacrylamide gel electrophoresis — (PAGE)

For the DNA separation of very small fragments (below 150 bp and resolution of length differences below 5 bp) polyacrylamide gels were used (12 % concentration). The gel chamber used was the *Mini Protean 3* system from *Biorad, California*. The running buffer was TBE 1x, ladders and sample buffers used were the same as for agarose electrophoresis (see section 2.2.2.5).

The ingredients for the acrylamide gel were mixed in a small beaker with a magnetic stirrer. Immediately before casting the gel, APS and TEMED were added and mixed thoroughly for 20 seconds. After casting the gel between the glass plates, the gel was allowed to settle for >1 hour at room temperature.

Prepared gels could be kept between the glass plates, covered in foil, overnight in the refrigerator before running them the next day.

The gel separation was performed at a voltage of ~18 V/cm. Staining and detection was the same as with agarose electrophoresis.

12 % acrylamide gel for DNA electrophoresis	
acrylamide solution 40 %, 19:1 acrylamide:bis-acrylamide	4,5 mL
TBE 5×	3 mL
ddH_2O	7,5 mL
immediately before casting, add 95 µl 10 % APS and 9.5 µl TEMED	

2.2.2.7 DNA gel extraction from agarose gels

DNA Bands of interest were cut out from the agarose gels in non-symmetric shapes (only gels run with TAE buffer, not TBE). Only flanking DNA bands had been stained with ethidium bromide so that the bands of interest were never in contact with ethidium bromide which could cause reduced efficiencies in further working steps.

The DNA was then extracted using the *Nucleospin Extract II* kit from *Macherey-Nagel, Germany*, according to the manufacturer's instructions. The method is based on denaturation of the agarose by a chaotropic salt (guanidine thiocyanate) and selective binding of DNA to a silica membrane.

2.2.2.8 DNA restriction

For the enzymatic digestion of DNA, different enzymes from *New England Biolabs, USA* were used. A reaction mix containing DNA solution, 1/10 of 10 × corresponding NEB buffer and an experimentally determined amount of restriction enzyme was set up. The final volume was set as low as possible, depending on the volume of DNA solution, but not lower than 20 µl in order to avoid evaporation effects. The mixture was then incubated at the optimum enzyme temperature.

In case of digestions with more than one enzyme, the corresponding buffer was determined through enzyme buffer compatibility charts available from the manufacturer.

Each DNA restriction was verified by gel electrophoresis. Depending on further utilization of the DNA, either gel extractions were performed to isolate the DNA fragments, or a cleanup procedure was used to remove enzymes, salt and potentially small DNA fragments (using *Nucleospin Extract II* kit from *Macherey-Nagel, Germany*).

2.2.2.9 DNA ligation

DNA fragments were ligated with the *Quick Ligation Kit* from *New England Biolabs, USA*, following the manufacturer's instructions. Volumes were kept as low as possible in order to aim for final DNA concentrations of ~10 ng/ µl. Different Vector DNA and insert DNA molar ratios were used to determine the optimal conditions (usually in the range of ratios from 3:1 to 1:3).

In the course of this thesis, only DNA ligations with overhanging 5' or 3' ends were done, and the 5'-ends were phosphorylated, the 3'-ends had hydroxyl groups. For those ligations, the ligation time varied between 15 min and 2 hr and the ligations were done at 16 °C.

After ligation, DNA was purified with the *Nucleospin Extract II* kit from *Macherey-Nagel, Germany* to remove salts and PEG from the ligation kit and prepare the DNA for electroporation.

2.2.2.10 Electroporation of DNA into *E. coli* cells

DNA used for electroporation of cells had to be prepared to avoid traces of salt in the transformation mix, which could cause a short circuit during the transformation, with the resulting heat destroying the cells. The *Nucleospin Extract II* kit from *Macherey-Nagel, Germany* was used for that purpose.

40 µl of electrocompetent cells were mixed with 1–3 µl of DNA (various concentrations, between 1 and 500 ng), all the time kept cool on ice. The mix was incubated for 1 min. Then the suspension was pipetted into an ice-cooled 2 mm electroporation cuvette (*Eurogentec, Belgium*) and electroporated in an Easyject electroporation device (*EquiBio Ltd., USA*) , which uses the parameters of 2500 V, 25 µF and 200 Ohm. "Successful" electroporation (low deviation from the ideal pulse time of 5 ms) was monitored by the device and communicated by an acoustic signal.

Immediately after the electroporation, 1 mL of prewarmed (37 °C) SOC medium was added to the cells, and the cell suspension was thereafter transferred into a sterile 15 mL conical tube. Cells were incubated at 37°C, 160 rpm shaking for 60 min.

This protocol follows the principles of Dower et al. (1988), who optimized parameters for electroporation for a number of bacterial and eukaryotic cells.

In order to determine the number of transformants, dilution series were made and plated out on antibiotic selective agar, while the rest of the cell suspension was plated out on large selective antibiotic agar plates.

SOC medium	
Tryptone	20.0 g
Yeast extract	5.0 g
NaCl	0.5 g
0.25 M KCl	10 mL
2 M $MgCl_2$	10 mL
1 M glucose	20 mL
ddH_2O	fill up to 1000 mL
adjust pH to 7.0 if necessary, then autoclave	

2.2.2.11 Generation of a dsDNA fragment from a ssDNA oligonucleotide

While the oligo-primed complementary strand synthesis of ssDNA DNA and following amplification by PCR is well known and established, in the course of this work any bias by amplification (Iannolo et al., 1997) was undesired. Thus, only a fill-out reaction (from ssDNA to dsDNA) was utilized. A combination of the heat-stable enzyme *Phusion DNA polymerase* (as in the PCR reactions [see chapter 2.2.2.3], a DNA polymerase with 5'-3' proof-reading activity) and T4 DNA polymerase (*New England Biolabs*) was used. The reaction mix also contained the four nucleotide triphosphates.

The mix for the first step was prepared and the incubated at 98 °C for 60 sec, 53 °C for 4 min and finally at 72 °C for 1.5 min. The DNA was then purified using the *Nucleospin Extract II* kit from *Macherey-Nagel, Germany*, eluted in 50 µl finally.

The purified DNA from the first step was then used in the second step to fill-out the remaining single stranded "ends" that had shown not be completely filled up in the first step alone. The ingredients were pipetted together and incubated at 12 °C for 60 min. After another purification with the *Nucleospin Extract II* kit (for each 150 µl, one column of the kit was used) and elution in 2×50 µl, the dsDNA DNA was prepared for validation by gel electrophoresis (preferably PAGE, s. chapter 2.2.2.6).

1st step dsDNA DNA fill-out	
5×HF Phusion buffer	10 µl
dNTP solution (2 mM each)	2 µl
50 mM $MgCl_2$	1 µl
DMSO	1.5 µl
Phusion polymerase	2 µl
Template (10 µM)	5 µl
Primer (10 µM)	15 µl
H_2O	13.5 µl

2nd step dsDNA DNA fill-out	
Elution 1st step	100 µl
dNTP (2 mM each)	20 µl
NEB 10×restriction buffer "2"	30 µl
T4 DNA polymerase	8 µl (=24 u)
H_2O	142 µl

2.2.2.12 DNA sequencing

In order to sequence plasmid DNA, 10 mL test tubes were filled with 2 mL 2 YT (+ respective antibiotics for plasmid/cell type selection) and inoculated by using sterile toothpicks that had been stabbed into isolated colonies. The test tubes were incubated over night at 37 °C, 200 rpm, next day a mini-prep plasmid isolation was done from those cultures (s. chapter 2.2.2.1).

Alternatively, 96 well deep-well plates filled with the same medium were used for cultivation later to allow for a higher throughput. Plasmid isolation was then kindly provided by Tschong-Hun Im from the department Genome Analysis at the Helmholtz Centre for Infection Research. The protocol and materials used there were similar to the ones used in commercial kits like the "QIAprep 96 Turbo Miniprep Kit" (*Qiagen*).

The plasmid DNA was sequenced by the department Genome Analysis, using appropriate primers that were provided. The DNA sequence data was then provided in .abi file format (Applied Biosystems sequencing data file).

2.2.3 Work with proteins

2.2.3.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

For the separation of proteins obtained from phage preparations, SDS poly-acrylamide gel electrophoresis (SDS-PAGE) was performed.

The gel chamber system used was the *Mini Protean 3* system (*Biorad, California*) with glass plates creating 8 cm × 8 cm gels of 0.75 mm thickness. The gel used was a discontinuous SDS gel (Laemmli, 1970) consisting of a stacking gel and a separation gel.

The separation gel was cast first, until ca. 1.5 cm below the pocket forming comb. It was overlayed with isopropanol and given 30 min to solidify. The isopropanol was removed then and the stacking gel poured into the glass cassette, sealing with the pocket forming comb.

Protein samples were precipitated with acetone before a gel run in order to remove salt from the sample: 1 volume of acetone was added to the protein solution (here: phage preparation, ~5×10^{10} phage particles, also diluted, depending on experiment) in an Eppendorf tube, mixed and cooled at -20 °C for at least 1 hr. The sample was then centrifuged in a lab centrifuge for 15 min at 13.000 rpm, the supernatant removed and the pellet suspended in a 1:4 mixture of 5 × protein sample buffer and ddH_2O .

The proteins in sample buffer were heated to 95 °C for 5 min in order to denature the proteins and loaded into the gel pockets after a short cooldown. The molecular weight standard used was *PageRuler Prestained Protein Ladder* (*Fermentas, Canada*).

The gel was run at constant 10 mA until the bromophenol-blue dye front reached the top of the separating gel, then the current was changed to a constant 20 mA until sufficient separation was achieved.

Separation gel (10 %)	
acrylamide solution 40 %, 19:1 acrylamide:bis-acrylamide	2.5 mL
1 M Tris-HCl pH 8.8	2.5 mL
10 % SDS	0.1 mL
ddH_2O	4.8 mL
right before casting, add 80 µl 10 % APS and 8 µl TEMED	

Stacking gel (4.5 %)	
acrylamide solution 40 %, 19:1 acrylamide:bis-acrylamide	550 µL
1 M Tris-HCl pH 6.8	625 µL
10 % SDS	50 µL
ddH_2O	3.75 mL
right before casting, add 40 µl 10 % APS and 4 µl TEMED	

Gel buffer 5 ×	
Tris base	15 g
Glycine	72 g
SDS	5 g
ddH_2O	fill up to 1000 mL

Protein sample buffer 5 ×	
1 M Tris-HCl pH 6.8	0.5 mL
Glycerol	1.6 mL
10 % SDS	1.6 mL
Bromophenol-blue solution 0.05 % (w/v)	400 µL
β-mercaptoethanol	400 µL
ddH_2O	5.5 mL

2.2.3.2 Western blotting

For western blot protein transfer, the *Mini Trans-Blot* system (*Biorad, California*) was used. It is made for the electro-transfer of proteins from the 8 cm × 8 cm gels used in the *Mini Protean 3* system (s. section 2.2.3.1).

A sandwich was prepared, starting with a fiber pad, then a layer of *Whatman 3 mm paper* (*Whatman, England*), followed by a nitrocellulose membrane (*Invitrogen*) and the acrylamide gel. The next layer was Whatman 3 mm paper and another fiber pad. The whole sandwich was soaked in blotting buffer, and potential bubbles were pressed out to the sides. The sandwich was put into the blotting cassette which was then inserted into the blotting chamber, then the chamber was filled with 1 × blotting buffer.

The electro-transfer was performed for 1 hr at a constant current of 200 mA.

After the transfer, the membrane was rinsed with water, then fixed (and temporary stained) with a solution of 0.2 % Ponceau red in 3 % TCA (Trichloroacetic acid). A photo of the temporary staining could be taken, then the Ponceau red was washed out thoroughly with water first, then PBS. After those steps, the membrane could be used for an immunostaining with specific enzyme coupled antibodies.

Blotting buffer 10×	
Tris base	30.3 g
Glycine	144 g
ddH_2O	fill up to 1000 mL
for 1×blotting buffer, mix 100 mL 10×blotting buffer with 50 mL methanol and 850 mL ddH_2O	

2.2.3.3 Immunological detection of proteins on Western blots

The nitrocellulose membrane with the blotted proteins was blocked with blocking solution for at least 2 h or over night. The blocking was followed by an incubation with an appropriately diluted primary antibody (in blocking solution) for the detection of the proteins of interest for 1.5 h. The membrane was washed three times for five minutes each with T-PBS, then incubated with an appropriate horseradish peroxidase labeled secondary antibody (diluted in T-PBS) for 1.5 h. The membrane was washed then three times for five minutes with T-PBS, then three times with distilled water.

For the detection of the bound antibodies, the membrane was dried for a short time by putting its edge onto a tissue paper. The membrane was put into a 50 mL tube, then the freshly prepared ECL solution was added. The tube was slowly rotated for 1 min,

then the membrane was taken out and dried shortly again. The blot was put between two sheets of plastic foil and put into a closed cassette.

In a darkroom a sensitive film was exposed to the chemiluminescence emitted by the reaction and then developed and fixed to get an image.

Blocking solution	
Skimmed milk powder	10.0 g
PBS buffer	fill up to 200 mL

T(0.02 %)-PBS	
Tween-20	100 µl
PBS buffer	500 mL

ECL solution	
1.5 M Tris ph 8.8	333 µl
p-Coumaric acid	11 µl
Luminol (*Fluka 09253*)	25 µl
H_2O_2	1.5 µl

2.2.4 Preparation and handling of bacteriophages

2.2.4.1 Preparation of M13K07 helper-phage stock

A fresh overnight culture in 2 YT was inoculated from a single Top 10 F' colony grown on M 9 minimal agar. This overnight culture was used to inoculate a 20 mL 2 YT culture which was grown at 37 °C, 160 rpm up to an OD_{600} of 0.5. The temperature of 37 °C is crucial for the induction of F-pili formation in the cells (Novotny and Lavin, 1971); these pili are necessary for an efficient infection of the *E. coli* cells by the filamentous phage M13.

500 µl of that culture were infected with 1 µl of several dilutions of M13K07 bacteriophage (1×10^{11} pfu, *New England Biolabs, USA*) and incubated at 37 °C for 15 min.

3 mL of premelted 2 YT Top Agar (as 2 YT agar, just 0.6 % agar, premelted and kept at 50 °C) were added to each 501 µl of cells, mixed briefly and then immediately poured onto a 2 YT agar plate (containing tetracycline).

After hardening of the Top Agar, the plates were incubated overnight at 37 °C.

From a plate that showed easily distinguishable, yet separate phage plaques, a single plaque was picked using a sterile plastic pipette tip stabbing through the agar. Note: Very small or very large plaques were avoided in order not to pick phages that might have developed mutations.

This piece of agar was used to inoculate 20 mL 2 YT (containing tetracycline and kanamycin) and grown for 6 h at 37 °C, 160 rpm. From this culture, 1 L 2 YT (containing tetracycline and kanamycin) was inoculated and grown overnight at 37 °C, 160 rpm.

Next day, the cells were centrifuged (15 min at 8000 rpm, 4 °C, GS3 rotor). The supernatant was carefully decanted and centrifuged a second time in order to remove remaining bacterial cells. The supernatant was then mixed with 0.15 volume of PEG/NaCl and put on ice for at least 2 h. The phage were precipitated by centrifugation (90 min, 8000 rpm, 4 °C, GS3 rotor). The supernatant was decanted and remaining liquid removed after a second short centrifugation (5 min, 8000 rpm, 4 °C, GS3 rotor).

The pellet was taken up in 10 mL PBS and cleared by 15 min centrifugation at 11000×g in a lab centrifuge. NaN_3 was added to the supernatant to a final concentration of 0.02 % NaN_3.

The phage titer had to be determined, it was usually in the range of 1–5×10^{12}.

2.2.4.2 Determination of bacteriophage titers

To determine bacteriophage titers, a 20 mL culture (2 YT, tetracycline) was inoculated with 200 µl of Top 10 F' overnight culture and grown at 37 °C, 160 rpm until an OD_{600} of 0.5 was reached.

The phage solution to be titered was diluted in PBS (10^0-10^{10}-fold), and 100 µl of each dilution were mixed with 100 µl of the growing culture in a well on a microtiter plate. The plate was incubated at 37 °C for 60 min, then 20 µl from each sample were spotted on 2 YT agar plates containing appropriate antibiotics (tetracycline and kanamycin to determine M13K07 titers, tetracycline and ampicillin to determine phagemid titers).

The agar plates were incubated overnight at 37 °C, then the colonies in each spot were counted and multiplied with the respective dilution factors in order to calculate the cfu.

2.2.4.3 Packaging of phagemid DNA libraries into bacteriophage particles

After DNA modification work on the library, either during the creation of the library or after modifications of the library later on, a large number of competent Top 10 F'λ cells were transformed with the phagemid DNA by electroporation (see section 2.2.2.10)

After electroporation, the cells resuspended in 1 mL SOC medium were plated out on 2 YT agar plates (containing tetracycline and ampicillin) to harvest the phagemid clones. 10 mL 2 YT medium (containing tetracycline and ampicillin) was pipetted onto every plate, then the plates were left resting for 30 min. The cells were scraped off from he plates using a Drigalski spatula, mixing them thoroughly in this process, then 3 mL from each plate of the same transformation experiment was pipetted into a 50 mL tube.

The cell suspension in the tube was mixed by inverting the tube several times, and 2.5 mL were uses to inoculate 250 mL 2 YT medium (containing tetracycline and ampicillin). From the rest of the cell suspension, several glycerol stocks were prepared by mixing 1 mL of cell suspension with 0.5 mL of sterile glycerol.

The 250 mL culture was grown at 37 °C, 160 rpm until an OD_{600} of 0.5 was reached. 2.5×10^{11} cfu of M13K07 were added to the culture, then the culture was incubated at 37 °C for 30 min without agitation to allow for the phage to infect the cells after regeneration of F' pili on the cells. The culture was then grown for six more hours at 37 °C, 160 rpm.

The cells were centrifuged (15 min at 8000 rpm, 4 °C, GS3 rotor). The supernatant was carefully decanted and centrifuged a second time in order to remove remaining bacterial cells. The supernatant was then mixed with 0.15 volume of PEG/NaCl and put on ice for at least 1 h. The phage were precipitated by centrifugation (90 min, 8000 rpm, 4 °C, GS3 rotor). The supernatant was decanted and remaining liquid removed after a second short centrifugation (5 min, 8000 rpm, 4 °C, GS3 rotor).

The pellet was taken up in 2 mL PBS and cleared by 15 min centrifugation at 13000 rpm in a bench top centrifuge. NaN_3 was added to the supernatant to a final concentration of 0.02 %.

The helper phage as well as the phagemid titer was determined by cfu dilution plating with Top 10 F' λ cells (see section 2.2.4.2).

Cells obtained from individual electroporations were initially kept separately. Pooling was only carried out after appropriate quality control checks.

2.2.5 Affinity selection from phagemid libraries

2.2.5.1 Panning procedure

A standard protocols was used to enrich candidate clones with a specific affinity to an immobilized target from a (sub-)library of phagemid particles.

The target was immobilized and incubated with the phagemid library. This was followed by washing steps to remove unspecific binding, and afterwards bound phagemid particles were eluted. Those remaining phagemid particles were used to reinfect Top 10 F‘ or Top 10 F‘λ cells, which, after superinfection with M13K07 helper-phage, produced new phage that could be subjected to another round of panning.

Affinity selections were performed either in reaction tubes containing magnetic beads that bind the target, or on microtiter plates where the target is adsorbed to the specially treated surface of the microtiter wells.

For immobilization on magnetic beads, protein A dynabeads (*Dynabeads Protein A, Invitrogen, California*) or streptavidin dynabeads (*Dynabeads M-280 Streptavidin, Invitrogen, California*) were used. The protein A, which is covalently bound on the bead surface, allows the immobilization of antibody Fc fragments of many different IgG antibodies, while the streptavidin allows immobilization of biotinylated molecules. The beads

present a considerably larger surface area than the wells of microtiter plates, which are often used in similar experiments. When the beads should be held back from a solution, the reaction tube was put into a magnetic stand (*Dynal MPC, Invitrogen*).

In order to prepare protein A beads for a panning experiment, the following procedure was used: 40 µl of Dynabeads suspension per 5 µg of IgG antibody/IgG antibody fusion protein to be immobilized were pipetted into a reaction tube. After putting the tube into the magnetic rack, the supernatant was removed. The beads were washed twice with a 500 µl 0.1 M Na-phosphate pH 8. The supernatant was discarded every time.

The washed beads were resuspended in a volume of 0.1 M Na-phosphate pH 8 so that together with the target protein solution the original volume of the Dynabeads was reached. The target protein solution was then added (between 1 and 50 µg of target). The tube was incubated at RT for 30 min under continuous slow inversion to avoid settling of the beads. After this capturing process, the tube was put into the magnetic rack and the supernatant was removed. The beads were then resuspended and washed with a volume of 500 µl 0.1 M Na-phosphate pH 8 three times.

Using streptavidin beads for a panning experiment, the following procedure was used: 100 µl of Dynabeads suspension per 5 µg of biotinylated IgG antibody fusion protein to be immobilized were pipetted into a reaction tube. After putting the tube into the magnetic rack, the supernatant was removed. The beads were washed twice with a 500 µl PBS. The supernatant was discarded every time.

The washed beads were resuspended in a volume of PBS so that together with the target protein solution the original volume of the Dynabeads was reached. The target protein solution was then added (between 1 and 10 µg of target). The tube was incubated at RT for 30 min under continuous slow inversion to avoid settling of the beads. After this capturing process, the tube was put into the magnetic rack and the supernatant was removed. The beads were then resuspended and washed with a volume of 500 µl PBS three times.

For the panning, the prepared beads were incubated with 500 µl blocking buffer for 30 min under slow rocking. This pre-blocked the tube walls as well as the protein beads and reduced unspecific binding of the phage particles later.

In parallel, 250 µl of phagemid library solution were mixed with 750 µl blocking solution (this was for a first round of panning, in following rounds, always 50 µl phage solution were mixed with 450 µl of blocking buffer), and also incubated for 30 min at room temperature.

After the 30 min, the beads were washed 3×with 500 µl T-PBS, then the phage-blocking-buffer-mix was added to the beads and they were incubated for 2 h at RT under gentle shaking.

The supernatant was removed from the beads, then the beads were washed several times with T-PBS and blocking solution — 10 min T-PBS, 10 min blocking solution, 5 min T-PBS, 5 min blocking solution, 5 min T-PBS, 5 min blocking solution. Finally, the beads were suspended in 200 µl elution buffer and shaken at RT for 15 min. Then the supernatant, containing the eluted phages (that had not been washed away, but were now eluted, was pipetted out of the tube and mixed with 40 µl neutralization buffer. Those 240 µl now contained the eluted phages.

T(0.05 %)-PBS (T-PBS)	
Tween-20	500 µl
PBS buffer	1000 mL

Blocking buffer (2 % MP)	
T(0.05 %)-PBS (T-PBS)	100 mL
Skimmed Milk Powder	2 g

Elution buffer (0.1 M Glycine)	
Glycine	35.4 g
ddH_2O	ad 200 mL
adjust pH to 2.2	

Neutralization buffer (2 M Tris)	
Tris base	24.2 g
ddH_2O	ad 200 mL

2.2.5.2 Reinfection of *E. coli* cells with eluted phages

20 µl of the elution solution were used to determine the titer of the phages after elution, the remaining amount was used to infect 15 mL of a freshly grown OD_{600}=0.5 culture of Top 10 F'λ in 2YT medium. The cells were mixed with the phage thoroughly and then incubated at 37 °C for 60 min. The cells were then pelleted by centrifugation (10 min, 5000×g) and resuspended in 400 µl 2YT (containing ampicillin and tetracycline). This suspension was plated out on a 125 mm ⌀ 2YT agar plate (containing ampicillin and tetracycline) and incubated overnight at 37 °C.

2.2.5.3 Packaging of phagemid particles from reinfected cells

10 mL 2 YT medium (containing tetracycline and ampicillin) was pipetted onto every plate on which the reinfected cells had been grown, then the plates were left resting for 30 min. The cells were scraped off from he plates using a spatula, mixing them thoroughly in this process.

2 mL of that cell suspension were used to inoculate 200 mL 2 YT medium (containing tetracycline and ampicillin). From the rest of the cell suspension, several glycerol stocks were prepared by mixing 1 mL of cell suspension with 0.5 mL of sterile glycerol.

The 200 mL culture was grown at 37 °C, 160 rpm until an OD_{600} of 0.5 was reached. 2×10^{11} cfu of M13K07 were added to the culture, then the culture was incubated at 37 °C

for 30 min without agitation to allow for the phage to infect the cells after regeneration of F' pili on the cells. The culture was then grown for six more hours at 37 °C, 160 rpm.

The cells were centrifuged (15 min at 8000 rpm, 4 °C, GS3 rotor). The supernatant was carefully decanted and centrifuged a second time in order to remove remaining bacterial cells. The supernatant was then mixed with 0.15 volume of PEG/NaCl and put on ice for at least 1 h. The phage precipitate was harvested by centrifugation (90 min, 8000 rpm, 4 °C, GS3 rotor). The supernatant was decanted and remaining liquid removed after a second short centrifugation (5 min, 8000 rpm, 4 °C, GS3 rotor).

The pellet was taken up in 200 µl PBS and cleared by 15 min centrifugation at 13000 rpm in a bench top centrifuge. NaN_3 was added to the supernatant to a final concentration of 0.02 % NaN_3.

The helper phage as well as the phagemid titer was determined afterwards, the phage solution could be used now for another round of panning. After the second round of selection, single clones were picked from the overnight plates for characterization by DNA sequencing of the library sequence containing the variable cassettes.

2.2.5.4 Characterization of selected single clones after reinfection

When packaging phagemid particles as in chapter 2.2.5.3, the phage titer was determined as well. Single colonies from those titer determination agar plates were picked and cultivated as in chapter 2.2.2.12, so that the DNA sequence of eluted phagemid particles could be determined.

2.2.5.5 Biacore affinity measurements

The Biacore system is based on a surface plasmon resonance (SPR) to monitor the adsorption of molecules on a sensor chip. This optical technique measures changes in the refractive index of the medium very close to a metal surface. Such changes are directly proportional to the amount and molecular weight of the macromolecules bound to the surface at any time.

This system is designed for functional real-time characterization of the interaction of protein-protein, ligand-receptor and the interaction of low molecular weight molecules in general. While all the interactions are dynamic processes, the ability to measure the kinetic parameters of association and dissociation (k_{on} and k_{off}) precisely allows the understanding of biological processes as well as the determination of affinity values (K_D) of molecules for targets of interest.

The Biacore system consists of a sensor chip on which the ligand can be immobilized on a matrix composed of flexible dextran, a miniaturized fluidics cartridge for the transport

of analytes and reagents to the sensor chip, and a SPR detector (Jönsson et al., 1991; Malmqvist, 1993).

A laser source emits plane-polarized light which is aimed through a prism, onto a metal covered sensor chip (usually gold or silver) on which the targets of interest are immobilized. When the analyte binds to this target, the beam angle is altered, which can be measured so that the intensity of the interaction can be determined. These measurements are performed continuously so that a real time plot (sensorgram) of the beam angle difference, in arbitrary resonance units (RU), versus time can be obtained (Panayotou et al., 1993).

From this data, using different mathematic fitting algorithms, the association and dissociation values of the studied interaction can be determined.

The measurements were done at the university of Kassel, Germany, in the department of Biochemistry. Sonja Schweinsberg (University of Kassel, Germany) performed the experiments.

The instrument used for the interaction studies was a Biacore 3000 (*BIAcore, Uppsala, Sweden*).

The biotinylated ligands (peptides with a biotin molecule covalently coupled to the C-terminus), solved in PBS buffer (concentration: 5 µg/mL), were immobilized on streptavidin SA chips from Biacore, which have streptavidin coupled covalently to the dextran layer on the chip surface.

Before immobilization of the ligands, the four flow cells of the chip were rinsed three times with 1 M NaCl/50 mM NaOH for 1 min at a flow rate of 1 µL/min.

The ligand peptides were then immobilized at flow rates of 10 µL/min up to 130–250 RU each.

The analyte studied in regards to its interaction with the immobilized peptides was a recombinant chimera protein of the extracellular domain of human CD28 (amino acid residues 1–152, Aruffo and Seed, 1987) fused carboxy-terminally to the Fc region of human IgG1 via a polypeptide linker (source: *R&D systems, USA*), suspended in PBS.

The kinetic measurements with the analyte were performed at a flow rate of 30 µL/min with an association phase of 120 s.

Regeneration of the chip surface after addition of the analyte was attempted by rinsing the flow cells twice for one minute each at a flow rate of 30 µL/min, using 1.5 M NaCl, 3 M NaCl and 10 mM glycine (pH 3).

2.2.5.6 Phage ELISA

T(0.05 %)-PBS (T-PBS)	
Tween-20	500 µl
PBS buffer	1000 mL

Blocking buffer (2 % MP)	
T(0.05 %)-PBS (T-PBS)	100 mL
Skimmed Milk Powder	2 g

0.05 M Citric Acid	
Citrate monohydrate	1.05 g
ddH_2O	100 mL
Adjust pH to 4.0 with 10 N NaOH and filter (0.2 µm) for sterilization	

1 ×ABTS stock solution	
ABTS diammonium salt (2',2'-azino-bis (3-ethylbenz-thiazoline-6-sulphonic acid))	22 mg
0.05 M Citric Acid solution	100 mL
Filter (0.2 µm) for sterilization	

ABTS substrate solution	
H_2O_2, 30 %	36 µL
1 ×ABTS stock solution	21 mL
Prepare just immediately before use	

In order to compare the interaction between enriched clones and the target (relative affinity ranking of clones), phagemid particles were packaged from cultures cultivated from single colonies. These phagemid particles were submitted to a phage ELISA (Enzyme Linked Immuno Sorbent Assay) experiment.

The target protein of interest was coated on Nunc-MaxiSorp microtiter plates (*Nunc/Thermo Fisher Scientific, Roskilde, Denmark*) by incubating wells with each 100 µL of protein solution (10 µg/mL) overnight at 4 °C in a humidified container.

The target immobilized in the microtiter wells was a recombinant chimera protein of the extracellular domain of human CD28 (amino acid residues 1–152, Aruffo and Seed, 1987) fused carboxy-terminally to the Fc region of human IgG1 via a polypeptide linker (source: *R&D systems, USA*), suspended in PBS.

The supernatant was removed from the wells, which were washed then with T(0.05 %)-PBS. After the washing, the wells were filled in with the blocking buffer and incubated at room temperature for 1 hr. Uncoated wells were blocked as well to serve as background controls.

In separate Eppendorf tubes, phagemid particles were diluted in blocking buffer to a final concentration of 5×10^{10} particles/mL and incubated for 30 min to block any non-specific protein-protein interactions that may occur between phagemid particles and surrounding proteins.

After removal of the blocking buffer from the microtiter plate, 200 µL of the phagemid solution was added to the wells and incubated at room temperature for 2 hours.

After incubation with the phagemid particles, the wells were washed twice with T(0.05 %)-PBS and then incubated for 1 hr with 200µL HRP conjugated anti-M13 antibody (article code 27-9421-011, *GE Healthcare, United Kingdom*), diluted 1:5000 in blocking buffer.

The microtiter plate wells were then washed three times with T(0.05 %)-PBS, and 200 µL ABTS substrate solution were added to the wells. After 5 to 30 min, depending on the development of the green color, the absorbance at 410 nm was measured in a Tecan Sunrise ELISA-plate reader (*Tecan, Männedorf, Switzerland*).

3 Results and discussion

3.1 Library design

The CPL19YS-2 library was designed with several considerations in mind (see Figure 16):

- **Peptide length:** Overall length of 33 amino acids, with as many positions variable as possible. This length should give potential for the formation of secondary structures that increase the molecule's internal stability (internal β-sheet structures) or the ability to interact with target domains (in the case of α-helices or loop formation), thus increasing the potential to bind target domains specifically and with higher affinity. This is discussed in more detail in 1.4.
- **Four cassettes:** The library comprises four cassettes as so called "building blocks". These building blocks can be recombined by *Cosmix-plexing* as described in chapter 1.4.2 during every stage of work with the library.
- **Potential for disulphide bridged loops:** In order to support the forming of stabilizing loops, a certain potential for the occurrence of multiple cysteine amino acids at defined positions was implemented.

The CPL19YS-2 library was designed to be cloned into the pEPO 8 phagemid vector (see section 3.2.1.4). This fuses a leader peptide sequence to the library's N-terminus, providing secretion out of the bacterial cell after cleavage of the leader peptide. Additionally, a full-size pIII protein from M13 is fused to the library's C-terminus, leading to the presentation of the library encoded peptide when incorporated into a phage particle (see Figure 15).

3.1.1 Peptide length and amino acid composition

Most peptide libraries consist of 6 to 12mers. The reason for this is that such a peptide library can actually cover nearly completely its whole potential complexity.

A library of 7mers, using all 20 available amino acids, would already have a potential complexity of 20^7= ca. $1.28{\times}10^9$. The phage libraries with the highest complexity, produced on an industrial scale, are in the ~10^{10}-range (see Vaughan et al., 1996)).

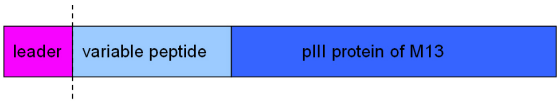


Figure 15: Schematic structure of the fusion protein encoded in the CPL19YS-2 library.
To the N-terminus of the library encoded peptide a leader peptide sequence is fused, providing secretion out of the *E.coli* cell while having the leader peptide cleaved off during secretion (Szardenings and Collins, 1990). The pIII protein of M13 is fused to the C-terminus of the library peptide via a linker sequence.

Although there are examples of short peptides possessing secondary structures, chances for structural features other than just extended linear or randomly coiled peptides are much higher with increased length (s. chapter 1.4).

As illustrated by the DNA sequence for peptides encoded in the CPL19YS-2 library in Figure 16, not every possible amino acid is available at each position (which would require NNB [with B=G,T or C] or NNK [with K=G or T] codons), but the repertoire at each position is limited. The reasons for this restriction were:

- **avoidance of stop codons:** the codons TAG, TAA and TGA were not supposed to exist in the CPL19YS-2 library at all, in order to prevent premature termination of translation even in the absence of suppressor tRNAs.

- **avoidance of additional restriction enzyme recognition sites:**
  The *Cosmix-Plexing* method used for recombination in the CPL19YS-2 library uses certain type IIs DNA restriction enzymes to cut the library on distinct positions that will be used for later re-ligation as well. Therefore it is essential for an efficient recombination process that all restriction enzymes used for the *Cosmix-Plexing* do recognize only *one* site per plasmid (a site inside the library region) so that an enzymatical digestion leads only to a liberalization of the plasmid rather than a fragmentation. It also helps in the last step of *Cosmix-Plexing*, where linear DNA fragments are re-constituted to circularized plasmids (see section 3.2.2.5), as re-circularization of plasmids is one of the most efficient ligation reactions (compared to the ligation of separate DNA fragments into a circular plasmid).

  The DNA triplets were chosen in a way that ensures that none of the restriction enzymes used (*Bce*AI, *Bsr*DI and *Bpm*I) will recognize positions inside the library other than the single predefined sites.

- **avoidance of the *HPQ* motif (histidine-proline-glutamine:** *HPQ* is an amino acid motif that is binding to streptavidin with a high affinity (Devlin et al., 1990). Streptavidin-coated surfaces or beads are often used to immobilize targets for use in a panning experiment, so it is undesirable to have the library peptides

binding strongly to the immobilization material rather than the target material. This measure reduces the background binding in the library that is often a problem in panning processes where streptavidin is used (Menendez and Scott, 2005).

- **avoidance of the *RGD* motif (arginine-glycine-aspartic acid:** *RGD* is a common motif present in many proteins that interact with integrins found on eukaryotic cells (Ruoslahti, 1996). Because it is such a small amino acid motif, the RGD site can be found in many random peptides, promoting cell adhesion. Analogous to the HPQ motif, a library not containing RGD could be used for screening targets presented on cell surfaces without enriching sequences that have an artificial RGD-based affinity to integrins on the cell surface without actually binding specifically to a presented target molecule. (Menendez and Scott, 2005).

- **absence of tryptophan in the library:** while tryptophan is an amino acid that can have a strong impact on a potential peptide structure, it is also very hydrophobic. In small peptides, a massed occurrence of tryptophan cannot be "shielded" by sequestering these hydrophobic residues in a "cage" of more hydrophilic amino acids. This can often lead to insoluble products (Riechmann and Winter, 2000), which are obviously undesirable for panning conditions in aqueous solutions. Another factor is that in many panning experiments, tryptophan-rich sequences were often preferentially enriched —- apparently not due to specific interactions with the target, but just due to hydrophobic "sticking" to the panning container (Adey et al., 1995); long aliphatic arms of arginine had also been noted in association with the same phenomenon. This target-unspecific binding should be reduced by removing tryptophan from the library.

- **restriction to just one of the acidic amino acids (aspartic and glutamic acid):** since those two amino acids have a rather similar structure and both contribute the negative charge of the COOH group, codons were selected in a way that only allows aspartic acid to occur within peptides of the CPL19YS-2 library.

- **absence of lysine and low frequency of arginine:** in order to avoid any additional recognition sites for the restriction enzymes used for recombination (see above), the sequence variety is limited. Due to this restriction, no lysine is encoded in the library sequence, and arginine is allowed only in 8 out of 25 positions. This contributes to positively charged amino acid residues being slightly underrepresented in the library and needs to be taken into account when analyzing enriched sequences.

- **coverage of variant combinations after recombination of selected variants:** even with the above mentioned limitations, the complexity of each of the four cassettes would still have been much higher than what would be possible to cover in the process of library creation (assuming that the library should cover at least most of its potential complexity per cassette). Therefore, as also common in other libraries (e.g. Dennis et al., 2002), the complexity of each *cassette* was adjusted to be below 1×10^8 by appropriate choice of degenerate codons.

  Even the complexities used in this design approach might still be too high, following what Clackson and Wells (1994) stated in their review. According to the authors, the actual number of clones required to ensure that all possible variants are present at least once in the primary library with a probability of 99 % is at least 5×greater than the potential complexity of a library (in our case, per DNA cassette).

  The necessary number of clones in the CPL19YS-2 library might be even higher, considering the potential bias in base composition of the synthetic oligonucleotides used for library creation (for the actual deviation from ideal distribution, see Table 2, p.70) and statistical variation (see Collins, 1997).

  Discounting sequence/codon bias, in order to cover the whole complexity within each cassette with a 99 % probability, the number of clones in a library would have to be approximately five times higher than the theoretical complexity, meaning a minimum of 2.75×10^8 clones — based on the complexity of 5.5×10^7 in the 3rd cassette).

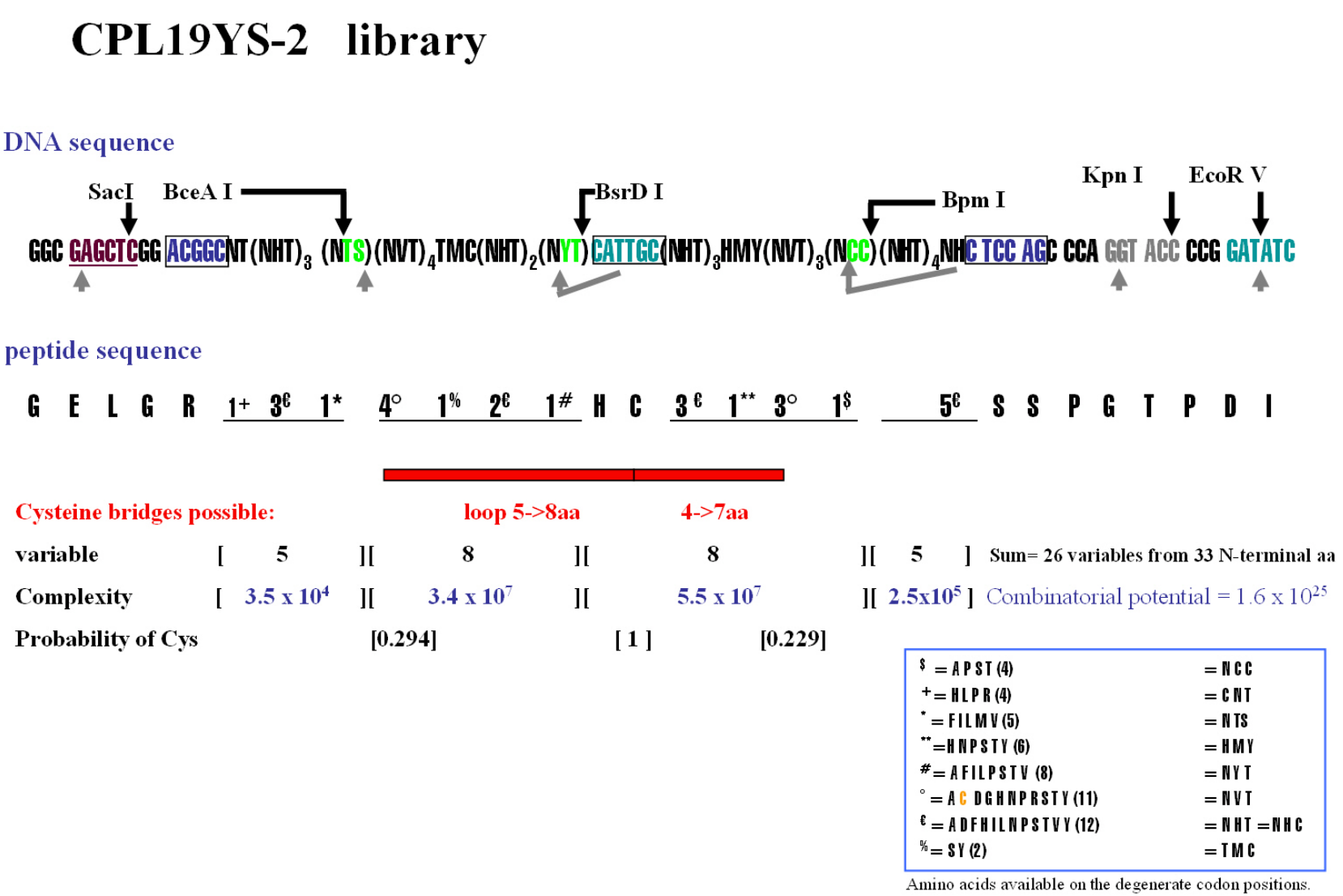


Figure 16: CPL19YS-2 library design. Library DNA sequence is shown on top, degenerated bases that allow for different nucleotides at their positions are H(A/C/T), M(A/C), S(G/C), V(G/A/C) and Y(C/T). The restriction sites for the enzymes *Bce*AI, BsrD I and *Bpm*I separate the four cassettes of the library, cutting positions are indicated by arrows. Below is the amino acid sequence with the variable positions shown as multiple occurrences of the special symbols $^{\$, +, *, **, \#, °, €}$ and $^{\%}$. The potential complexities due to the available amino acids are stated below the cassettes in blue. Potential for cysteine per cassette and possible disulfide bridge positions are indicated. The availability of amino acid at each library position is described in further detail in Figure 17.

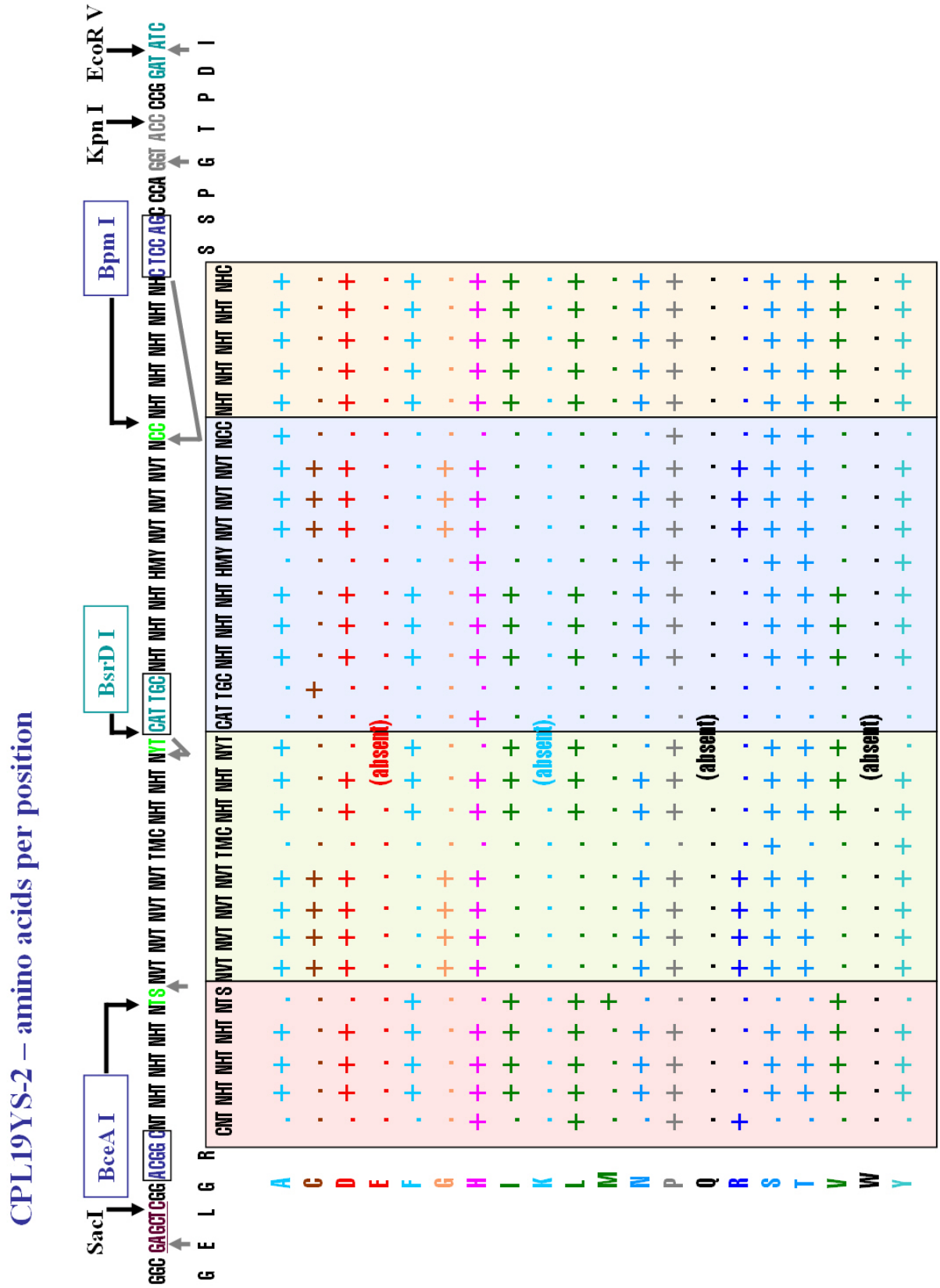


Figure 17: CPL19YS-2 library, amino acid availability per position.
The four colored boxes indicate the four cassettes of the library. A '+' indicates that an amino acid is available at this position, while a '·' indicates absence at that position. Amino acids E (glutamic acid), K (lysine), Q(glutamine) and W (tryptophane) are not available on the variable positions of the CPL19YS-2 library.

3.1.2 Cassette structure

While most of the sequence is degenerate (8 different types of degenerate codons were used), there are certain fixed positions in the library. Among those are the recognition sites for the three different type IIs restriction enzymes mentioned above (see e.g Szybalski et al., 1991, for a review of type IIs restriction enzymes).

Type IIs restriction enzymes cleave at a fixed distance adjacent to their recognition sequence. They recognize a sequence that is continuous and asymmetric. This feature makes it possible to position the (fixed) recognition sequence *outside* of the library encoding sequence, while still being able to cleave at a fixed position *within* the library sequence.

The positions at which the restriction enzymes cleave (highlighted in green in Figure 16) were chosen in such a way that, after a cleavage, the sticky-ends are non-palindromic. This ensures that during subsequent ligation of the cleavage products only an ordered recombination is possible (as required for *Cosmix-plexing*, section 1.4.2), rather than head-to-head/tail-to-tail ligations. All the cohesive-end sites are also specific to allow for a correctly ordered directed ligation-mediated recombination even if cleavage is carried out simultaneously with a combination of type IIs restriction enzymes.

3.1.3 Disulfide bridges

It has been shown that constrained sequences or peptides often possess advantages in affinity selections over linear peptides (see Binz and Plückthun, 2005).

One method to achieve this is to bracket a peptide sequence by a pair of flanking cysteines, which, when oxidized to a disulfide linkage, result in the displayed peptide being presented to the target as a looped structure. This method has been found to yield positive results (e.g. by McLafferty et al., 1993), and is commonly used, even in commercial libraries, such as the *Ph.D.-C7C* phage display library from *New England Biolabs* (presenting random heptamers, flanked by cysteines), or the *versabody* libraries from *Amunix* (*California, USA*), which consist of 15–60mers possessing additional rigidity due to multiple disulfide bridges (>15 % overall cysteine content).

The CPL19YS-2 library was designed to offer a multitude of a least partially constrained, highly variable three-dimensional structures, rather than being restricted to restrained peptides of fixed architecture. This was achieved by allowing for the occurrence of multiple cysteines within one peptide (positions with potential for presence of cysteine are marked by a red bar in Figure 16). This design element has been used in several other libraries before, for example by McConnell et al. (1996).

Whereas the central cysteine is fixed, the *NVT* codons to the left and to the right of the central cysteine have the *potential* to code for a cysteine. Assuming a Poisson

distribution, the probability for the occurrence of a peptide with just one cysteine (the central one) are 54.4 %, while the probability for one additional cysteine (thus having two cysteines and the chance to form disulfide bridges) is 38.8 %. The remaining 6.8 % represent the chance to have three or more cysteines in a peptide.

3.2 Library creation

3.2.1 The phagemid vector

The library was created by inserting synthetic, highly variable DNA fragments into a new DNA vector pEPO 8, based on the pROCOS 4/7 phagemid vector (provided by *Cosmix GmbH, Braunschweig, Germany*), which is derived from the pSKAN 8 vector (Röttgen and Collins, 1995).

The pROCOS 4/7 vector contained a library (15mer peptide library pCPL3, unpublished data, Michael Mersmann, John Collins), fused N-terminally to the full-length minor coat protein pIII of the M13 bacteriophage and fused C-terminally to a secretion-providing leader peptide. The pCPL3 library was replaced with the library created in the course of this thesis.

The properties of the pROCOS 4/7 vector are largely identical to the modified pEPO 8 phagemid vector, and are explained in detail in section 3.2.1.4.

3.2.1.1 Removal of restriction sites

In a first step, the pROCOS 4/7 vector had to be modified since it contained several restriction sites which were required to be unique in the final vector in order to allow recombination using selected type IIs restriction enzymes.

The DNA type IIs restriction enzymes used for recombination of the CPL19YS-2 library are *Bce*AI, *Bsr*DI and *Bpm*I. The original pROCOS 4/7 vector contained seven recognition sites for these restriction enzymes (see Figure 18).

The restriction sites were removed by mutagenesis PCR (see Figure 19) to generate a modified DNA fragment (no longer containing the unwanted restriction site) that was then cloned into the pROCOS 4/7 vector replacing the original part with the unwanted type IIs restriction site.

For every site to be mutated, plasmid DNA from the pROCOS 4/7 vector was used as a template for two separate PCR reactions. Hereby, two DNA fragments (of 200 - 800 bp length) were generated; one containing the sequence upstream of the site to be mutated, ranging up to a position of a DNA restriction site utilizable for cloning, the other fragment covering the DNA sequence downstream of the mutation site, including a restriction site suitable for cloning as well.

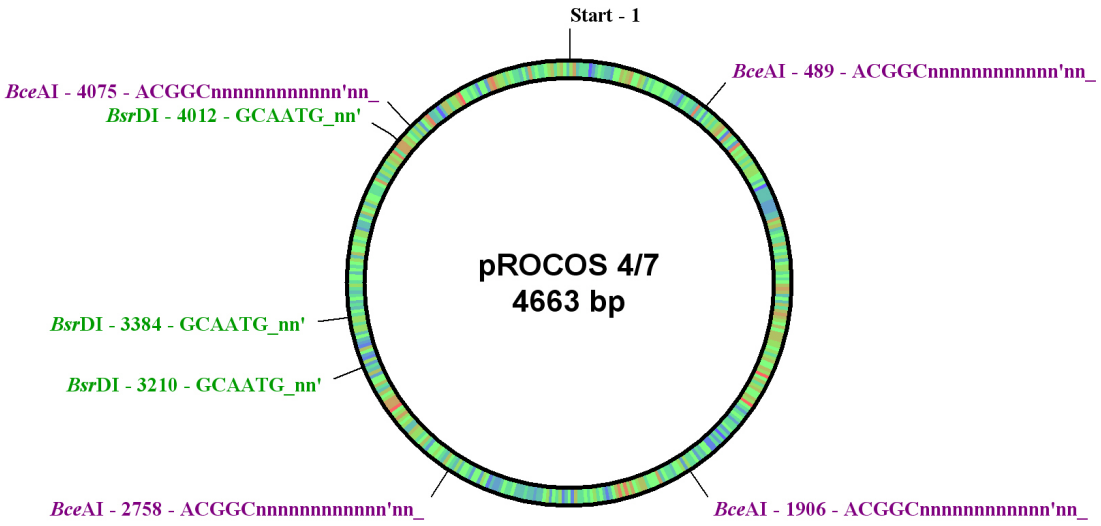


Figure 18: Phagemid plasmid vector pROCOS 4/7 with seven (unwanted) restriction enzyme recognition sites and their positions given as distance in base pairs relative to base pair position 1. There is no unwanted recognition site for *Bpm*I.

Both fragments overlapped (20–25 bp) in the sequence region containing the restriction site to be mutated; the primers used to mediate amplification in this region contained a point mutation that removed the restriction site in both PCR DNA fragments.

After a cleanup step (excision of the amplification product bands and subsequent extraction after agarose gel electrophoresis, in order to remove the last traces of remaining vector DNA), both separate PCR products were mixed together and used as template DNA for a further PCR. The flanking primers for this second round of PCR were specifically designed to yield a larger PCR amplicon containing the desired mutation, thus removing the original unwanted restriction site. Due to the choice of flanking primers, this PCR amplicon contained restriction enzyme recognition sites on both ends that allowed cloning into the pROCOS 4/7 vector.

These "corrected" fragments where subsequently cloned into the pROCOS 4/7 vector, using appropriate restriction enzymes and restriction sites as close to the site of mutation as possible, in order to minimize undesired mutations in other regions. In principle, it is not unlikely that such mutations occur during PCR amplification (ca. 1 base change per 1000 bp). The resulting vector was therefore always checked for successful mutation by digestion with the DNA restriction enzyme whose recognition site had been mutated. Only after verification of the successful restriction site modification, the vector was used as starting material for the next round of PCR-mediated site-specific mutation.

This strategy was used to eliminate the seven unwanted recognition sites shown in Figure 18. Unfortunately, during the final quality control it became apparent that an-

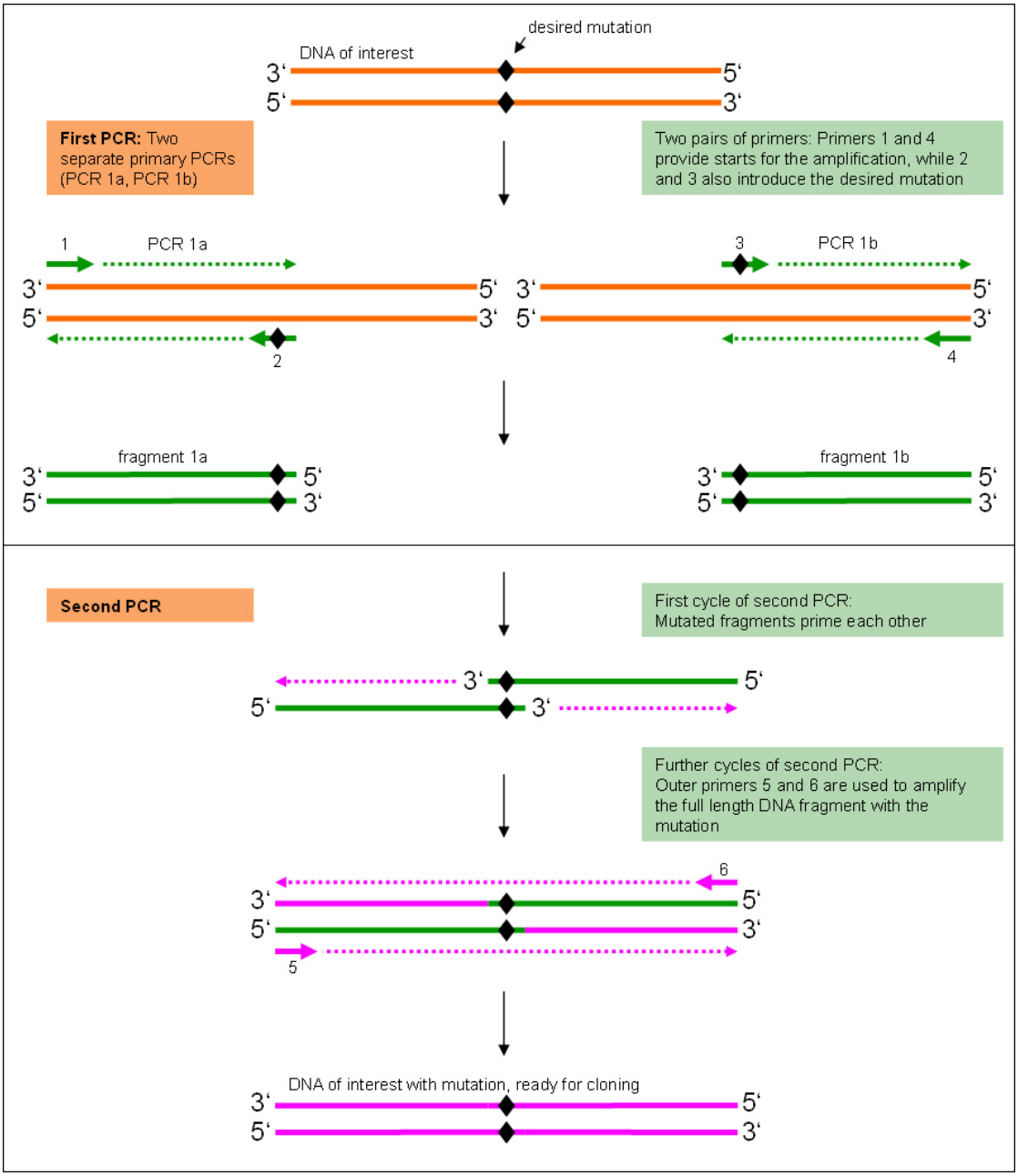


Figure 19: PCR mutagenesis.
Introduction of a mutation into a DNA sequence using two subsequent PCR reactions with specifically designed primers.

other recognition site for *Bsr*DI had been accidently introduced at position 4109 (most probably in the process of eliminating the *Bce*AI recognition site at position 4075). This site was therefore re-mutated into the original sequence of the pROCOS 4/7 vector, and this final product was designated pEPO 8.

3.2.1.2 Preparation of the library insertion site

The pEPO 8 vector contained a sequence of 76 bp length between the restriction sites chosen for library insertion (enzymes Sac I and Eco RV — see Figure 20).

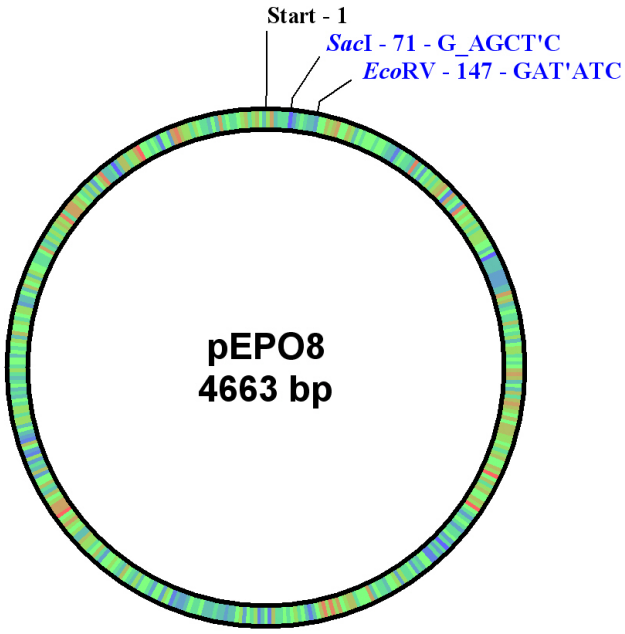


Figure 20: library insertion site for pEPO8

Thus, a digestion with those two enzymes would result in two fragments (4587 bp and 76 bp) that could not easily be distinguished from fragments that would results from only partial digestion — 4587 bp vs. 4663 bp.

It was, however, highly desirable to ensure that the completely double-digested vector would be clearly separated and easily distinguishable from the uncleaved and partially cleaved DNA during subsequent clean up gel electrophoresis. In order to achieve this, in an intermediate step a "stuffer" fragment was introduced between the *Sac*I and the *Eco*RV site of pEPO 8.

A stuffer fragment (1745 bp in size, obtained from Lambda DNA *Sac*I and *Eco*RV digestion) was cloned between the Sac I and Eco RV sites of pEPO 8. The resulting vector was then designated pEPO 8-stuffer. A digestion of this vector with the two restriction enzymes *Sac*I and *Eco*RV now resulted in easily separatable bands (4587 bp + 1745 bp for correct double digestion vs. 6408 bp for partial cleavage products).

3.2.1.3 Validating the vector sequence

After the introduction of the stuffer fragment, plasmid DNA from pEPO 8-stuffer was digested with *Bce*AI and *Bsr*DI to check if any unwanted restriction sites had been introduced or arisen by mutation. Since no further unwanted restriction sites outside of the stuffer fragment were found, the plasmid pEPO 8-stuffer was sequenced using various different primers in order to cover the whole sequence. Additionally, a large part of the vector DNA was sequenced forward and backward to reduce sequencing errors as much as possible. A few deviations from the sequence of the pROCOS vector, provided by the *Cosmix GmbH* (*Braunschweig, Germany*), were found during sequencing and noted in a corrected vector map.

These sequence deviations did not confer any functional alterations to the plasmid vector with respect to packaging, antibiotic resistance, promoters or DNA replication. All these factors were checked extensively to ensure full for functionality of the final library vector.

3.2.1.4 The pEPO 8 phagemid vector

The final form of the pEPO 8 phagemid vector is shown in Figure 21.

The leader peptide sequence Omp_L of the "outer membrane"-protein Omp A is fused to the first amino acid of the CPL19YS-2 library, thereby ensuring production, secretion and correct cleavage of the leader peptide (Szardenings and Collins, 1990). This in turn results in secretion of the fusion protein, while the leader peptide sequence (MKKTA-IAIAVALAGFATVAQA), which is fused N-terminally to the library insert, is cleaved off.

Within the pEPO 8 phagemid vector, the pIII-CPL19YS-2(library) fusion protein is expressed under the control of the p_L promoter from the bacteriophage *Lambda*. This promoter is strongly repressed by the λ cI protein that is produced in λ lysogenic *E. coli* strains. By using this promoter, it is ensured that the ratio of hybrid to wild type pIII in the phagemid particles (every particle presenting five copies of pIII) produced upon superinfection with the helper-phage M13K07 is approximately 1:10, i.e. one hybrid protein per two phagemid particles (Röttgen and Collins, 1995). This leads mainly to the desired monovalent presentation which is advantageous during selection for high affinity ligands (Lowman et al., 1991).

The pEPO 8 vector contains an ampicillin resistance gene (*bla*, β-lactamase) to allow selection for plasmid uptake, an Col E1 origin of replication responsible for plasmid replication in *E. coli* (ensuring a multicopy plasmid by relaxed replication) and the M13 phage replication/packaging origin *ori fd*.

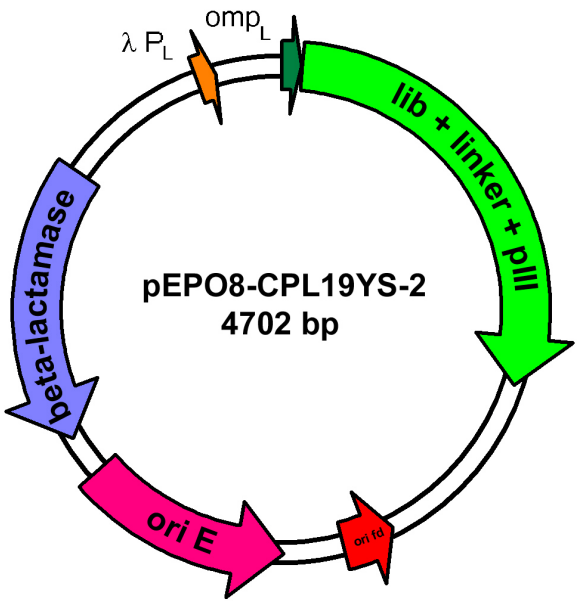


Figure 21: Phagemid plasmid vector pEPO 8, with inserted library CPL19YS-2 (section 3.2.2)
The following abbreviations are used: λ P_L: major leftward promoter of phage λ; omp_L: leader peptide signal sequence derived from the *E. coli* outer membrane protein Omp A; pIII: gene encoding the minor coat protein pIII of the bacteriophage M13K07; bla: β-lactamase gene, conferring ampicillin resistance; ori fd: origin of replication for single stranded DNA of the bacteriophage fd; ori E: origin of replication from the plasmid Col E1. A library DNA sequence, including a linker sequence, has been inserted in front of the pIII coding region.

3.2.2 Generation of the library

3.2.2.1 Preparation of linearized vector DNA

Large amounts of pEPO 8 stuffer plasmid DNA were isolated and enzymatically digested with Sac I and Eco RV. The DNA was separated on an agarose gel until good single band separation was achieved. Only the linearized, desired vector fragment of 4587 bp size (without the stuffer fragment of 1745 bp, see 3.2.1.2) was extracted from the agarose gel.

3.2.2.2 Generation of library dsDNA

The library encoding DNA oligonucleotides were ordered from the company *Biospring* (*Frankfurt am Main, Germany*). The single stranded DNA was filled up to dsDNA by using the method described in chapter 2.2.2.11, and afterwards enzymatically digested with Sac I and Eco RV in preparation for the primary library preparation (see Figure 22).

The ssDNA oligonucleotide used as "backbone" was designated CPL19YS-2 BAS, the oligonucleotide used for the fill-out reaction was JCOSTART3B-3 (see section 2.1.5.1).

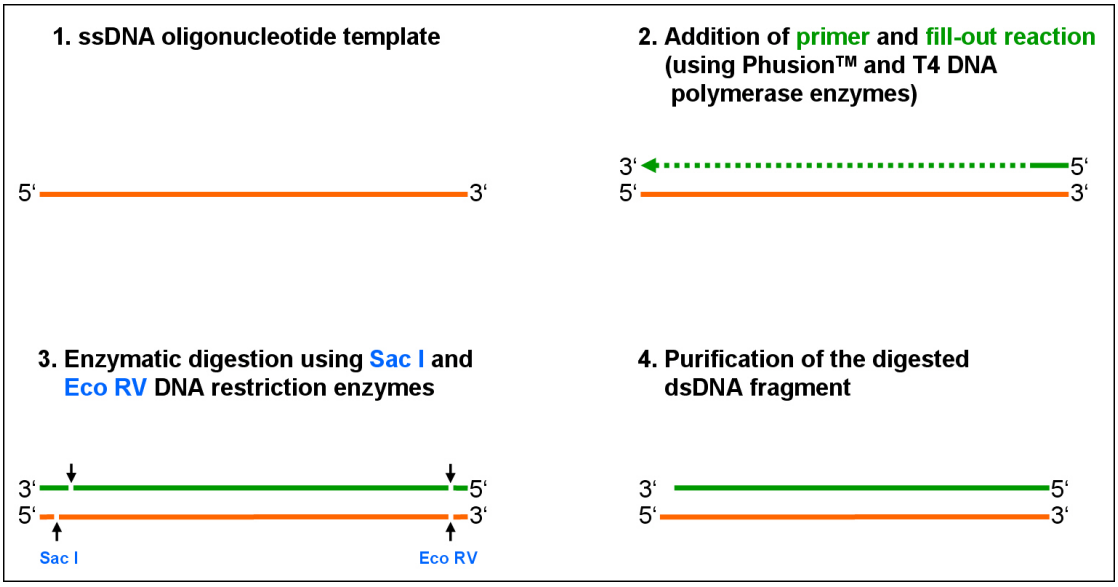


Figure 22: Preparation of the synthetic DNA library insert

3.2.2.3 Ligation of vector and library DNA

After validation of correct fragment sizes, the library dsDNA fragment and the large Sac I-Eco RV pEPO 8 vector fragment were ligated (see section 2.2.2.9) and subsequently transformed into Top 10 F' cells (see section 2.2.2.10). The amount of DNA used for electroporation was as high as 1–2 µg of DNA per electroporation step sometimes in order

to generate a high number of transformants (up to 1×10^7 transformants), compared to the recommended amount of 50 ng DNA/20 µl for optimal efficiency. There were about thirty transformations made, resulting in ca. 1.8×10^8 transformants total, which is also assumed to be the total complexity of the primary library.

Further work in our lab, conducted by Simone Nieswandt, increased the complexity of the primary library from 1.8×10^8 to about 4.2×10^8, using the Cosmix Plexing technique described in section 3.2.2.5 and the recombination sites *Bpm*I and *Bsr*DI.

This latter library was used in panning experiments conducted on the target CD28 (section 3.3.3) and the NS 3-4A + NS2-3 serin proteases of the hepatitis C virus. Subsequently the library has also been used for panning on various antibodies.

3.2.2.4 Evaluation of primary library constitution

After each separate electroporation batch, single clones containing the plasmid (indicated by growth on ampicillin antibiotic agar) were picked and sequenced (s. chapter 2.2.2.12).

The first part of the analysis was to confirm that each clone contained a single library insert, in correct orientation and open reading frame without additional mutations (insertions, deletions, frame shifts). The percentage of clones not conforming to these criteria was determined.

The second part of the analysis addressed was the statistical frequency of each single nucleotide at each position in reference to the expected values.

From DNA sequencing using the primers JCOFSQ (TTC TAC AAC TTG CTT GGA TT, binding 5' of the library sequence at vector position 4575, reading downstream) and JCORSQ (TCC AGA CGT TAG TAA ATG AA, binding behind (3' to) the library sequence at vector position 218, reading upstream) it was concluded that 10 from 132 sequenced clones did not meet the desired criteria. In 2 of them no insert was detected, the rest had various insertions or deletions. This corresponded to there being some 7.6 % "defective" clones in the library. The other 122 clones were of the correct sequence type and could be analyzed in their nucleotide composition. This frequency distribution was compared to the expected nucleotide sequence according to the library design (fig. 16, chapter 3.1).

As can be seen in Table 2, the deviations from the statistically expected nucleotide distributions were similar for each nucleotide, rather than deviating maybe dependent on codon usage. If the latter had been the case, one could assume selective pressure due to expression of the encoded peptides, but since the variations almost always had the same tendencies, it can be concluded that these differences originate in the synthesis process of the template DNA oligonucleotides, e.g. efficiency of chemical conjugation during oligonucleotide template synthesis.

In summary, the frequency of adenine is much less frequent than expected (ca. between 1/3 and 1/2 lower), while guanine is found with about 1/3–1/2 higher frequency than expected. Cytosine frequency deviation is statistically insignificant and inside the expectation range, while thymine frequency is slightly over-represented.

Averaging the frequency deviations in Table 2 and considering the design in section 3.1, the distribution of amino acids for the most common DNA triplets NYT and NVT (the most common triplets, on 19 of 25 variable positions) would result in reduction to ~40 % of the original value for threonine, to ~60 % for histidine, to ~50 % for isoleucine and to ~30 % for asparagine. The frequencies of occurrence for other amino acids were increased, as for alanine (increased to ~140 %), cysteine (increased to ~185 %, but only in the seven NVT triplets), phenylalanine (increased to ~140 %), valine (increased to ~150 %), glycine (increased to ~200 %) and arginine (increased to ~125 %).

In summary, this deviation would result in more hydrophobic peptides than originally devised. This has to be considered when analyzing enriched peptides.

The increase in cysteine frequency also increases the chances of more than one cysteine per encoded peptide. While the original design predicted 54.4 % of the clones with just the central cysteine, 38.8 % with two cysteines and 6.8 % of the clones with more than two cysteines, the 122 sequenced clones showed 33 sequences with just one cysteine (27.0 %), 59 clones with two cysteines (48.4 %) and 30 clones with three or four cysteines (24.6 %). This needs to be taken into account when analyzing peptides, as clones with two or more cysteines are much more frequent than originally intended.

Table 2: Nucleotide distribution for each position in 122 clones sequenced from the CPL19YS-2 library. The first % value for each nucleotide gives the frequency with which the respective nucleotide actually occurs at a certain position, the second value gives the deviation in percentage points by which the distribution in the library at the given position actually deviates from the "ideal" frequency. A blue number indicates a higher frequency than statistically expected, a red number a lower frequency.

Position	A	deviation	G	deviation	C	deviation	T	deviation
1	0%	0%	0%	0%	100%	0%	0%	0%
2	7%	-18%	39%	14%	18%	-7%	35%	10%
3	0%	0%	0%	0%	0%	0%	100%	0%
4	14%	-11%	35%	10%	19%	-6%	30%	5%
5	27%	-6%	0%	0%	30%	-3%	42%	9%
6	0%	0%	0%	0%	0%	0%	100%	0%
7	14%	-11%	29%	4%	22%	-3%	33%	8%
8	22%	-11%	0%	0%	38%	5%	38%	5%
9	0%	0%	0%	0%	0%	0%	100%	0%
10	9%	-16%	38%	13%	20%	-5%	31%	6%

continued on next page

Position	A	deviation	G	deviation	C	deviation	T	deviation
11	23%	-10%	0%	0%	31%	-2%	44%	11%
12	0%	0%	0%	0%	0%	0%	100%	0%
13	7%	-18%	34%	9%	18%	-7%	39%	14%
14	0%	0%	0%	0%	0%	0%	100%	0%
15	0%	0%	60%	10%	39%	-11%	0%	0%
16	8%	-17%	34%	9%	20%	-5%	36%	11%
17	27%	-6%	40%	7%	31%	-2%	0%	0%
18	0%	0%	0%	0%	0%	0%	100%	0%
19	14%	-11%	34%	9%	13%	-12%	36%	11%
20	21%	-12%	52%	19%	26%	-7%	0%	0%
21	0%	0%	0%	0%	0%	0%	100%	0%
22	16%	-9%	38%	13%	13%	-12%	31%	6%
23	17%	-16%	63%	30%	19%	-14%	0%	0%
24	0%	0%	0%	0%	0%	0%	100%	0%
25	13%	-12%	36%	11%	13%	-12%	36%	11%
26	17%	-16%	59%	26%	23%	-10%	0%	0%
27	0%	0%	0%	0%	0%	0%	100%	0%
28	0%	0%	0%	0%	0%	0%	100%	0%
29	45%	-5%	0%	0%	55%	5%	0%	0%
30	0%	0%	0%	0%	100%	0%	0%	0%
31	13%	-12%	40%	15%	13%	-12%	32%	7%
32	20%	-13%	0%	0%	29%	-4%	50%	17%
33	0%	0%	0%	0%	0%	0%	100%	0%
34	8%	-17%	38%	13%	23%	-2%	29%	4%
35	17%	-16%	0%	0%	28%	-5%	54%	21%
36	0%	0%	0%	0%	0%	0%	100%	0%
37	10%	-15%	47%	22%	18%	-7%	22%	-3%
38	0%	0%	0%	0%	34%	-16%	65%	15%
39	0%	0%	0%	0%	0%	0%	100%	0%
40	0%	0%	0%	0%	100%	0%	0%	0%
41	100%	0%	0%	0%	0%	0%	0%	0%
42	0%	0%	0%	0%	0%	0%	100%	0%
43	0%	0%	0%	0%	0%	0%	100%	0%
44	0%	0%	100%	0%	0%	0%	0%	0%
45	0%	0%	0%	0%	100%	0%	0%	0%
46	9%	-16%	39%	14%	22%	-3%	28%	3%
47	18%	-15%	0%	0%	42%	9%	38%	5%
48	0%	0%	0%	0%	0%	0%	100%	0%
49	11%	-14%	42%	17%	17%	-8%	28%	3%

continued on next page

Position	A	deviation	G	deviation	C	deviation	T	deviation
50	22%	-11%	0%	0%	35%	2%	42%	9%
51	0%	0%	0%	0%	0%	0%	100%	0%
52	11%	-14%	39%	14%	18%	-7%	31%	6%
53	27%	-6%	0%	0%	23%	-10%	49%	16%
54	0%	0%	0%	0%	0%	0%	100%	0%
55	14%	-19%	0%	0%	31%	-2%	53%	20%
56	36%	-14%	0%	0%	63%	13%	0%	0%
57	0%	0%	0%	0%	29%	-21%	70%	20%
58	14%	-11%	38%	13%	16%	-9%	30%	5%
59	19%	-14%	48%	15%	31%	-2%	0%	0%
60	0%	0%	0%	0%	0%	0%	100%	0%
61	9%	-16%	37%	12%	16%	-9%	36%	11%
62	18%	-15%	47%	14%	34%	1%	0%	0%
63	0%	0%	0%	0%	0%	0%	100%	0%
64	12%	-13%	39%	14%	22%	-3%	25%	0%
65	20%	-13%	51%	18%	27%	-6%	0%	0%
66	0%	0%	0%	0%	0%	0%	100%	0%
67	10%	-15%	39%	14%	26%	1%	23%	-2%
68	0%	0%	0%	0%	100%	0%	0%	0%
69	0%	0%	0%	0%	100%	0%	0%	0%
70	13%	-12%	37%	12%	19%	-6%	29%	4%
71	25%	-8%	0%	0%	35%	2%	39%	6%
72	0%	0%	0%	0%	0%	0%	100%	0%
73	10%	-15%	45%	20%	15%	-10%	27%	2%
74	20%	-13%	0%	0%	27%	-6%	51%	18%
75	0%	0%	0%	0%	0%	0%	100%	0%
76	10%	-15%	37%	12%	21%	-4%	30%	5%
77	15%	-18%	0%	0%	32%	-1%	51%	18%
78	0%	0%	0%	0%	0%	0%	100%	0%
79	9%	-16%	40%	15%	17%	-8%	31%	6%
80	21%	-12%	0%	0%	29%	-4%	49%	16%
81	0%	0%	0%	0%	0%	0%	100%	0%
82	9%	-16%	38%	13%	25%	0%	26%	1%
83	18%	-15%	0%	0%	31%	-2%	50%	17%
84	0%	0%	0%	0%	100%	0%	0%	0%

3.2.2.5 Evaluation of the Cosmix-plexing principle

With the primary library being cloned, the Cosmix-plexing feature had to be evaluated in function. Using this method, it is possible to increase the variety of the primary library by an ordered recombination between the cassettes. This recombination can generate new sequences every time, since the number of all *possible* cassette combinations in the CPL19YS-2 library is 1.6×10^{25}.

It is also possible to recombine selected sub-populations of clones either with themselves, another sub-population or the complete library. This can help to analyze the effect of certain sequence motifs on the panning process if those sequences are either kept or replaced by sequences from other cassettes. One can also use Cosmix-Plexing to generate new sequence parts "around" or in addition to an motif of interest. While a motif could have been enriched due to a positive effect on affinity, the neighboring sequences could still have potential for improvement, which could be possibly assessed by replacing the flanking cassettes with cassettes from the original library, adding variety that can be screened again.

The CPL19YS-2 library was designed to allow for the Cosmix-Plexing method of recombination, therefore it needed to be shown that the recombination of the library cassettes was possible and the procedure working without any bias which could affect screening experiments.

For reasons of verifiability, a limited repertoire of just four different clones from the library was selected for the recombination experiments. Recombination within such a limited repertoire allows to check if the expected combinations can be obtained in equal frequencies.

Using the restriction enzyme *Bsr*DI, the library region was cut in the middle and then recombined as shown schematically (Figure 23).

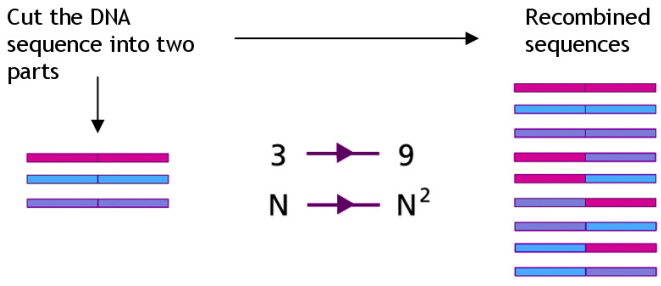


Figure 23: Principle of recombination and generation of variety from a limited repertoire of variants

Cosmix-plexing was performed using plasmid DNA isolated from the four different clones. DNA from the four variants was mixed in equal amounts. The principle of recombination can be seen in Figure 24.

The plasmid DNA mixture could be digested using any one of the three possible type IIs restriction enzymes, as shown in Figure 16 (here *Bsr*DI was used, cutting in the middle position between cassettes 2 and 3). After purification, the linearized vectors were ligated at DNA concentrations as high as possible (>200 µg/mL), forming concatemers. The specific nucleotides forming the "sticky ends" allowed only oriented ligation, no formation of head-to-head ligations.

The DNA concatemers were digested with a second restriction enzyme (*Bgl*I, cutting only once in the vector sequence) to generate linearized vector fragments that combined two different library cassettes now.

After digestion, the DNA was diluted to a low DNA concentration (<20 µg/mL) and ligated to make an internal ring closure the most frequent ligation product. The library was transformed into Top 10F'λ cells by electroporation and could also be packaged into phagemid format.

24 clones from the transformation were picked and sequenced (s. Figure 25). One of them had a frame shift mutation (meaning 4.2 % mutated variants, well below the level of degeneration in the library found before).

The 23 remaining sequenced clones showed an even distribution of the different library DNA fragment from the four original clones, which confirmed that there was no bias in the process of recombination. 14 different sequences were obtained from the 23 clones picked, which was a good yield considering that 16 different combinations of the original four sequences were possible. Sequencing of more clones would have shown the remaining two combinations in all likelihood.

The recombination process was shown to be reliable, without introducing any bias, degenerating library quality and allowed for an *ordered* recombination of cassettes. Recombination was also verified as working utilizing the other two restriction/recombination sites (*Bce*AI and *Bpm*I, data not shown here).

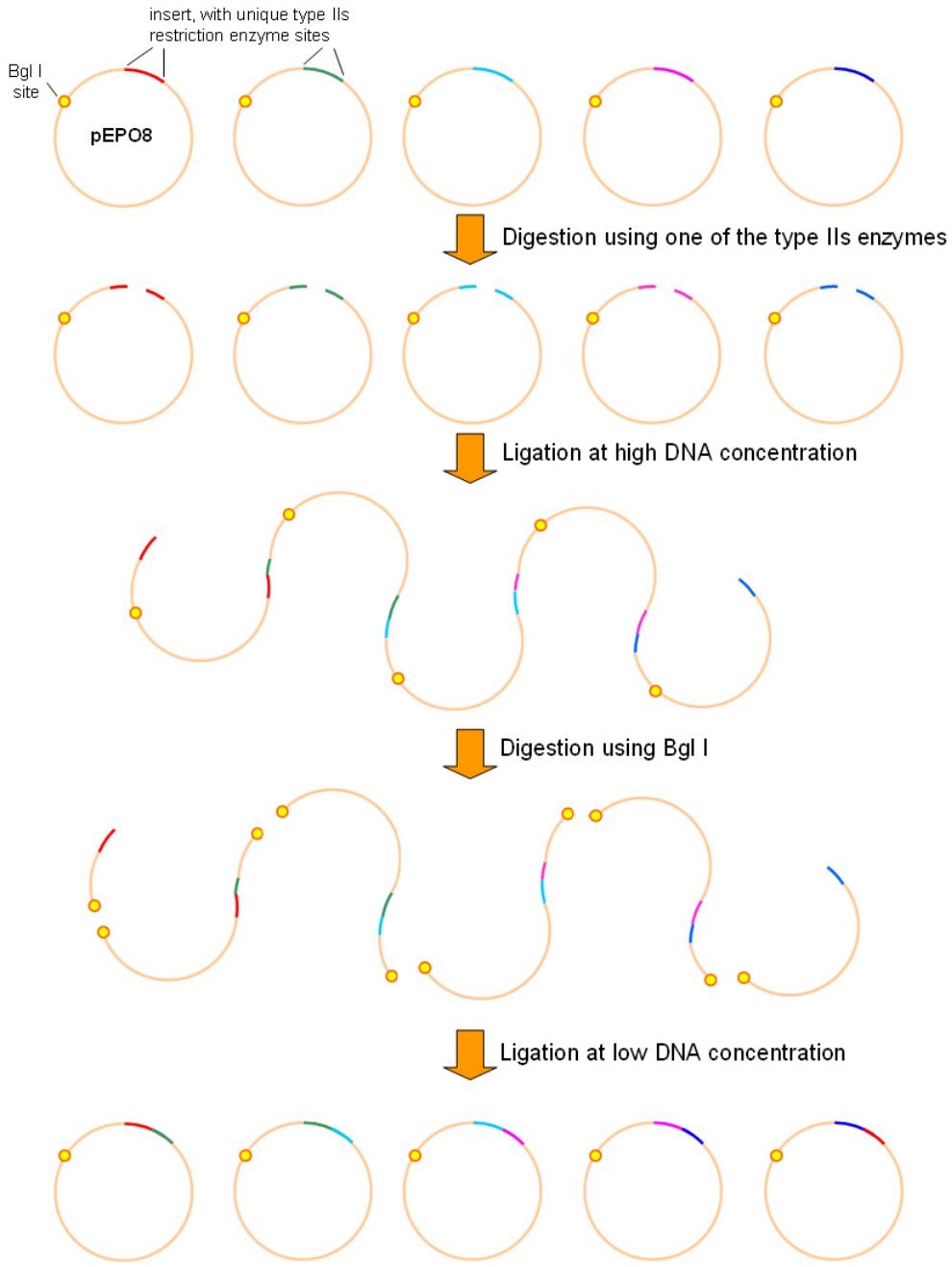


Figure 24: General method of Cosmix-plexing
The starting material consists of pEPO 8 phagemids carrying the library insert. These phagemids can either be from the original primary CPL19YS-2 library, or from eluted phage that have already been subjected to panning.
The phagemids are linearized using one of the three type IIs restriction enzymes cutting inside the library cassette region (*Bce*AI, *Bsr*DI and *Bpm*I) and then religated at high DNA concentration to form concatemers. The concatemers are resolved as single vectors by digestion with *Bgl*I, and ligated at low DNA concentration to achieve ring closure as monomers. The recombined phagemid vectors can be electroporated, packaged, and used for panning.

Figure 25: Evaluation of Cosmix-plexing: The four clones from the library, B43, B45, B46, B47 were recombined with each other at the middle position (*Bsr*DI site, indicated by the arrow). The four clones were chosen that produced the same "sticky-ends" at positions 50–51 (AA) to allow recombination between all four candidates.
After recombination, the four starting clones were shown together with 23 recombined clones. Between the new clones, 14 different sequences were found after recombination, from 16 different possible clones.

3.3 Evaluation of the CPL19YS-2 library

3.3.1 Evaluation of peptide display

With the DNA sequence quality of the phagemid in plasmid format validated, it was still necessary to validate the display of the fusion peptide on the phage surface itself. As mentioned already in chapter 1.3, there are different modes of expression if the peptide is fused to the pIII phage protein. Using the pSKAN 8 expression vector from which the plasmid pEPO 8 is derived, a ratio of about 1:10 for hybrid protein to wildtype pIII protein had been found in the past (Röttgen and Collins, 1995). This ratio had been determined by western blotting phage preparations and immunofluorescence staining of the pIII protein. The size difference between the native pIII and the fusion protein (fusion protein + pIII) was used to separate those two proteins and allowed a comparison of the band intensities.

Attempts to do so with the CPL19YS-2 library had been unsuccessful, it had not been possible to visually identify a band differing in size from the native pIII protein. However, successful enrichment during panning *was* demonstrated possible. It was therefore concluded that just the size difference between the native pIII (running in a gel at about an apparent size of 60 kDa) and the fusion protein (60 kDa + [114 Da (average aa mass) ×33= ~3.7 kDa]= about 63–64 kDa) was not large enough, or other structural factors made a separation impossible.

In order to remedy this effect, in our lab a partial deletion mutant of pIII was created (John Collins, unpublished). Following ideas from Holliger and Riechmann (1997), a deletion of the domains D1 and D2 of pIII (see Figure 26; see section 1.3.1.1 for explanation of the domains' functions) in the library vector was created in our lab by Simone Nieswandt and Gökhan Haslak. This fd-D3 mutant was used to display a peptide from the CPL19YS-2 library. Due to the reduction in size (D3 domain: ~24 kDa + (peptide mass: 114 Da ×33 = ~3.7 kDa) = ~28 kDa) compared to the wild type pIII (~60 kDa apparent size), a second band would be easier to separate on a Western blot.

PEG/NaCl precipitated phagemid particles from either full-length pIII fusion protein (type fd-D123 – packaged phagemid pEPO 8-Do1-19)) or partially deleted pIII fusion protein (type fd-D3 – packaged phagemid pEPO 8-Do1-19-D3) were examined in a Western blot with monoclonal anti-pIII antibody (*MobiTec, Göttingen, Germany*). The antibody binds to the -A T D Y G A A I D G F- epitope that lies in the D3 domain of pIII and can therefore be used to detect the truncated pIII protein as well. Figure 27 shows additional bands around 26 kDa in the fd-D3 type that could not be found in the fd-D123 phagemid preparation. While the bands between 34 and 43 kDa only appeared at high protein concentrations (also in the fd-D3 preparation), the 26 kDa bands were only detected

in the fd-D3 phagemid alone. From this, it can be concluded that those bands show the truncated pIII fusion protein. The multiple bands were assumed to be degradation products.

Concluding from the serial dilutions in Figure 27 the ratio of truncated pIII fusion protein to wildtype pIII in the fd-D3 type was determined to be ~5–15 % (confirmed by two more experiments, also with a type fd-D23 construct, data not shown here). These ca. 10 % were also expected from the data in Röttgen and Collins (1995) and were confirmed to be valid with the modified phagemid vector pEPO 8.

These data show both that the display of the fusion protein on the phage coat protein was taking place and gives in addition information about the valency of the fusion protein per phage particle.

With ~10 % of the pIII proteins being fusion proteins, there would be about one fusion protein per ten wildtype pIII proteins (meaning one fusion protein per two phagemid particles, with each particle presenting five pIII proteins). This would result in most phagemid particles presenting either none or just one fusion protein (assuming a Poisson distribution and 10 % probability of the fusion protein: 0 hybrid molecules – 60.7 %; monovalent – 30.4 %; divalent – 7.5 %; trivalent – 1.25 %; 4 hybrids – 0.156 %; 5 hybrids – 0.016 %; i.e., ca. 39.3 % of the phage particles are presenting the fusion protein, according to Röttgen and Collins, 1995). This leads to prevalently monovalent display which is important for the selection of high affinity ligands from a library.

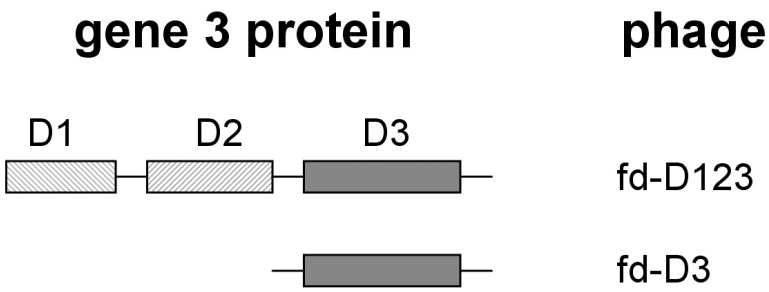


Figure 26: Schematic representation of the gene3 protein (pIII) in phage fd constructs. Boxes indicate the presence of the domains g3p-D1, g3p- D2, and/or g3p-D3. Lines indicate the glycine-rich linkers between g3p-D1, g3p-D2, and g3p-D3 and the C-terminal membrane anchor, respectively.

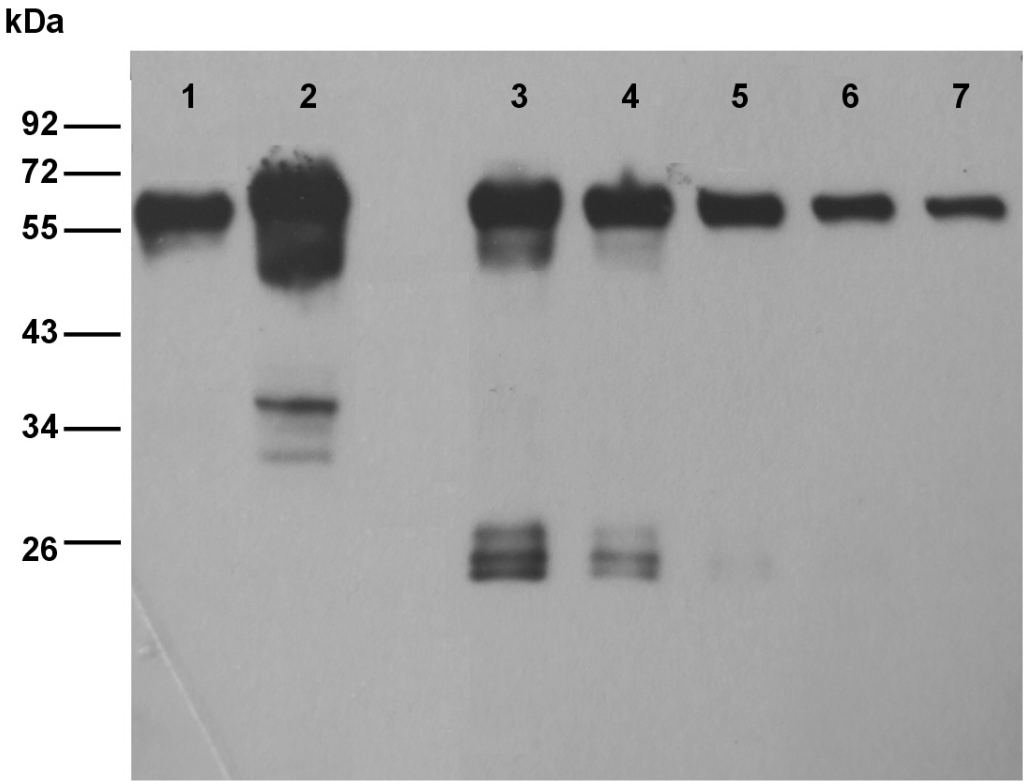


Figure 27: Western blot of phage preparations showing pIII and pIII- fusion proteins. Lanes: 1, pEPO 8-Do1-19 phagemid (wildtype pIII and pIII fusion protein); 2, threefold amount of phagemid as in lane 1; 3–7, pEPO 8-Do1-19-D3 phagemid (wildtype pIII and truncated pIII+fusion protein) in serial 2-fold dilutions from 3 to 7.
Phagemid was packaged from *E. coli Top 10 F'*λ cells using M13K07 helper-phage.
Phagemid particles were precipitated by PEG/NaCl precipitation. The protein separation was performed in a discontinuous SDS gel (4.5 % stacking gel, 10 % separation gel, after Laemmli, 1970), followed by a western blot transfer onto nitrocellulose. Protein pIII was detected by a monoclonal anti-pIII antibody, for staining a secondary peroxidase-labeled antibody was used.

3.3.2 Evaluation panning of the CPL19YS-2 library on the model target anti-p53 Ab-6 (Clone Do-1)

The antibody anti-p53 Ab-6 (Clone Do-1 – source: *Labvision, USA*) was selected as a model target to determine appropriate panning conditions for the CPL19YS-2 library and to validate that library peptides are being displayed. The antibody is a monoclonal mouse antibody targeting the p53 tumor suppressor protein. The p53 protein is expressed in a wide variety of tissue types and regulates the cell cycle.

The binding specificity of the antibody had been well studied in the past (for example Stephen et al., 1995), with affinity having been shown for the motif (F/Y)SDLx(R/K)(L/M), with x being no specific amino acid.

It was expected that similar motifs would be enriched during panning the CPL19YS-2 library on this antibody if the peptides were being efficiently presented on the surface of the phagemid.

3.3.2.1 Panning conditions

The panning was performed as described in chapter 2.2.5.1, the panning/washing conditions (and thus the panning stringency) were not changed over the different rounds of panning due to the monitored enrichment. The amount of target immobilized for each panning was 5 µg of p53 Ab-6 in the first three rounds of panning, reduced to 2.5 µg in the 4th round and to 1.25 µg in the fifth round.

There were two independent pannings done in parallel (P#1 and P#2), as well as a background control panning with no antibody immobilized on the protein A beads (PB). The background control panning was stopped after the third round because all isolated clones were found to be mutants, containing stop codons or deletions.

3.3.2.2 Enrichment of eluted phages

If an affinity selection panning is successful, over the course of several rounds the recovered clones will possess improved binding properties. This results usually in an increased ratio of phage recovered after the panning (phage output) in relation to the number of phage incubated with the target (phage input) should go up. This increase in the ratio can usually be seen as a positive indication for a successful panning process, i.e. enrichment of higher affinity ligands.

An enrichment in this panning experiment was only detectable in the fifth round in one of the two separate panning experiments (P#2), as can be seen in Table 3. While the output:input ratio often goes down in the first few rounds of a panning experiment, which was also the case here, the number of eluted phage goes up only in the case of P#2 in the 5th panning round (see also Figure 28). The very high amount of recovered phage in the first round of panning on the background (PB) seems to be due to a mistake in the experimental work, since it could not be reproduced later under similar conditions.

	P #1	P #2	PB
1R phage input	2.5×10^{11}	2.5×10^{11}	2.5×10^{11}
1R phage output	4×10^{6}	4×10^{6}	4×10^{7}
1R output:input	$\mathbf{1.6 \times 10^{-5}}$	$\mathbf{1.6 \times 10^{-5}}$	$\mathbf{1.6 \times 10^{-4}}$
2R phage input	2.5×10^{10}	3×10^{10}	2×10^{10}
2R phage output	5×10^{5}	5×10^{5}	1.6×10^{3}
2R output:input	$\mathbf{2 \times 10^{-5}}$	$\mathbf{1.7 \times 10^{-5}}$	$\mathbf{8 \times 10^{-8}}$
3R phage input	2×10^{10}	3×10^{10}	1.2×10^{10}
3R phage output	2×10^{4}	2×10^{4}	1×10^{4}
3R output:input	$\mathbf{1 \times 10^{-6}}$	$\mathbf{6.7 \times 10^{-7}}$	$\mathbf{8.3 \times 10^{-7}}$
4R phage input	2.5×10^{10}	3×10^{10}	
4R phage output	8.3×10^{4}	1.3×10^{4}	
4R output:input	$\mathbf{3.3 \times 10^{-6}}$	$\mathbf{4.3 \times 10^{-7}}$	
5R phage input	3×10^{10}	2.5×10^{10}	
5R phage output	4×10^{4}	4×10^{6}	
5R output:input	$\mathbf{1.3 \times 10^{-6}}$	$\mathbf{1.6 \times 10^{-4}}$	

Table 3: Development of phage titer over five rounds on panning the CPL19YS-2 library on p53 Ab-6.
P#1 and P#2 are the two separate pannings with p53 Ab-6 as target, PB is a panning on just the beads used for immobilization to check the background enrichment. 1R to 5R indicates the round of panning, input gives the number of phage incubated with the target protein, while output shows the amount of eluted phage after washing steps.
An increase of the output:input ratio from one round to another usually indicates a successful enrichment of clones presenting ligands with high affinities.

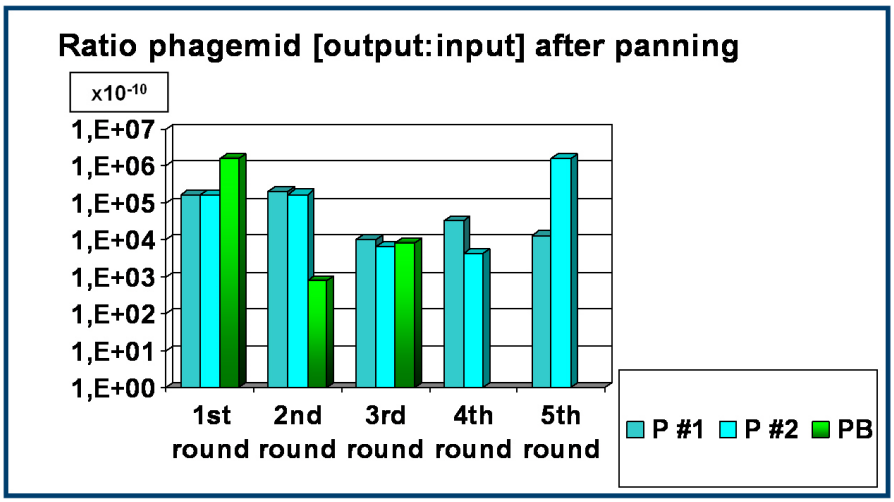


Figure 28: Output:Input ratio for different rounds of panning the CPL19YS-2 library on the p53 Ab-6 target

3.3.2.3 Analysis of isolated phages

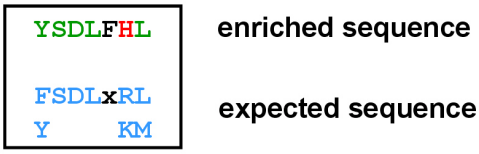
In addition to measuring the output:input ratio of eluted phage after each round of panning, single clones were also picked from each population of eluted clones. For the background (PB), there were only clones from the 3rd round analyzed (6 total), while for the two parallel pannings (P #1 and P #2) clones from the 4th and the 5th round were sequenced (6 clones for each round and panning).

While the DNA sequences from the PB clones showed 4 stop codon mutations and 2 mutations where no library insert was detectable at all (thus, 100 % of the clones were defective), the 24 sequenced clones from the two pannings gave just one single stop codon mutation (4.2 % mutations).

Figure 29: Sequences of single clones isolated after 4/5 rounds of panning on p53 Ab-6, panning P#1
The variable part is flanked by the aa sequence ELGR to the left and by SSPG to the right. HC in the middle is also invariable. 4R indicates 4th panning round, 5R means 5th panning round. 1 out of 6 clones in the 5th round was a mutant (only five sequences therefore after the 5th round). Even after five rounds of panning, there was no enrichment of a specific clone or consensus sequence detected.

Figure 30: Sequences of single clones isolated after 4/5 rounds of panning on p53 Ab-6, panning P#2
The variable part is flanked by the aa sequence ELGR to the left and by SSPG to the right. HC in the middle is also invariable. 4R indicates 4th panning round, 5R means 5th panning round. Already after four rounds of panning, a dominant sequence appeared, after five panning rounds there is only one enriched specific sequence left in all six sequenced clones.

The sequences from panning P #1 in Figure 29 showed no detectable enrichment of sequences similar in the variable part of the library. In the second panning, P #2 (see Figure 30), however, a specifically enriched sequence could be detected. Since there were no different sequences, but just one single recurring sequence, no consensus sequence could be concluded. But the similarity between the sequence element YSDLFHL with the expected sequence (F/Y)SDLx(R/K)(L/M) is obvious:



The only difference between the result and the expectation is thus the histidine instead of the arginine/lysine. This can be explained by the non-availability of R or K at this position in the library (see chapter 3.1), and taking into consideration that histidine can be seen as similar to those amino acids in hydrophilicity and under the pH of the panning buffer having a positive charge. Histidine can thus be considered an acceptable substitution for arginine or lysine.

3.3.2.4 Conclusions from panning on the model target anti-p53 Ab-6

The most important conclusion to be drawn from this panning experiment was that the panning & selection process worked — a proof of principle. While a negative result would have given no enriched clones, or only clones containing non-displaying mutations as was found in the negative control panning, a successful enrichment was achieved, confirmed by comparison with an already known sequence.

Presentation of the fusion protein was confirmed independently by antibody mediated detection of the fusion protein (chapter 3.3.1), it was still necessary to confirm the actual process of panning to be working.

It is noted that only a single clone had been found to be enriched as opposed to an enrichment of clones with a similar or identical binding sequence (consensus motif) with different flanking sequences (=variants with a consensus motif). This expected "richness" of clones would result from panning a library of sufficient complexity, i.e., which would hold a range of sequences containing such sequences. This result came along with the observation that only in one of two experiments an enrichment was observed *at all*.

There are three possible explanations for this result — either the complexity of the library was much lower than expected, or the washing conditions in the first rounds of panning were too stringent, so that not all the ligands showing affinity to the target were retained and eluted. The last explanation would be that there were too few panning

rounds to bring out other clones and that there were not enough clones sequenced to determine that no other clones have been enriched.

Since there were no other indications of a much lower library complexity than expected, and one panning did not yield any enrichment at all, it was more likely that there were limitations in how many of the ligands expressing higher affinities were kept after the first round of panning.

The first round of panning usually is the most important one with regard to accessing the whole complexity of a library, since clones from a complex library usually appear only in a low number of ($<10^4$) copies (see section 1.3.3). If the washing conditions are too stringent, coupled with a low number of copies per clone, it can happen that only a few of the high affinity ligands are actually retained and could be eluted for the consequent panning rounds, resulting in less or no high affinity ligands enriched in the process, even though those ligands appeared in the original library.

Considering that only a single clone, no variations of it, had been isolated, and that only one of two experiments yielded an enriched clone, it seems highly improbable that the general complexity of the library was too low. If a low library complexity would be the reason for the absence of variations on the expected binding sequence, both experiments should have yielded the same resulting clone — which was not the case. This supports the hypothesis that the results were originating from too stringent washing in the first round. It was also confirmed by further work with the CPL19YS-2 library conducted in our laboratory by Juliane Lindner and Jonas Kügler (Helmholtz Centre for Infection Research, see section 3.3.5).

As a consequence, in the later pannings the washing times were shortened only for the first panning round, the other parameters remaining constant as in the general guidelines in chapter 2.2.5.1. The conditions used are given in the respective experimental description.

3.3.3 Test evaluation panning on CD28

CD28 is a member of the immunoglobulin superfamily of molecules (IgSF). It is a homodimer of two polypeptides containing a single V-like domain with short transmembrane and cytoplasmic regions. It serves as a co-signalling molecule for T cell activation through binding to its cognate counter-receptors B7.1 (CD80) and B7.2 (CD86), expressed on antigen presenting cells. The costimulation is important for the proliferation of T cells and induction of an antigen specific immune response Linsley and Ledbetter (1993).

There are bacterial superantigens which trigger an excessive Th1-cytokine response leading to toxic shock. These superantigens (e.g. Staphylococcal Enterotoxin B, SEB, produced by *Staphylococcus aureus*), are indicated (Arad et al., 2009, submitted) to inter-

act directly with CD28. Rather than interacting indirectly via the CD80/B70 receptors expressed on antigen presenting cells, they lead to massive induction of Th1 cytokines (Arad et al., 2001) which include interleukin-2 (IL-2), interferon-gamma (IFN-γ), and tumor necrosis factor β (TNF-β or LT-α)(see Hackett and Stevens, 1993).

Due to this indicated direct interaction of superantigens with CD28, the interaction sites of CD28 are of interest to characterize, since blocking of CD28-superantigen interaction sites could help to alleviate the symptoms of a toxic shock.

In the screening of the CPL19YS-2 library, no specific domain of CD28 was targeted, the library was screened for any clones that exhibit affinity to CD28.

The CD28 molecule used as a target in the panning experiments was from two different sources to allow two different modes of immobilization:

- A recombinant chimera of the extracellular domain of human CD28 (amino acid residues 1–152, Aruffo and Seed, 1987) fused carboxy-terminally to the Fc region of human IgG1 via a polypeptide linker. Source: *R&D systems, USA*

- A recombinant fusion protein of human CD28 (extracellular domain of CD28, amino acid residues 1–134, Aruffo and Seed, 1987) and the Fc portion of mouse IgG2a (232 aa), which was biotinylated after expression and purification.
  Source: *Ancell, USA*

3.3.3.1 Panning conditions

The panning was performed as described in chapter 2.2.5.1, with the exception that the washing stringency in the first round was reduced in order to retain more phage (following the conclusions in chapter 3.3.2.4). Therefore, in the first panning round the washing steps consisted of rocking incubation for 10 min T-PBS, 10 min blocking solution and 5 min T-PBS before doing the elution. In the following rounds, the complete washing procedure was used and not changed over the course of the panning experiments.

The panning/washing conditions (and thus the panning stringency) were not changed over the different rounds of panning (rounds 2 to 4) due to the monitored enrichment. The amount of target immobilized was reduced for each round of panning, as shown in Figure 31. After the second round of panning, the amplified phages were not just panned further on the protein A immobilized CD28, but a fraction was panned against the streptavidin bound biotinylated CD28. This alternation of the material used as immobilization surface should prevent any enrichment of clones which specifically bind to either streptavidin or protein A.

This has previously been recognized as a problem (Menendez and Scott, 2005), where for instance HPQ-containing sequences had been enriched on streptavidin. It should

further be noted that the motif HPQ should not occur in the variable region of the CPL19YS-2 library.

Two independent pannings were performed (P #1 and P #2), as well as two background control pannings with no target/antibody immobilized on the protein A beads (PB#1 and PB#2).

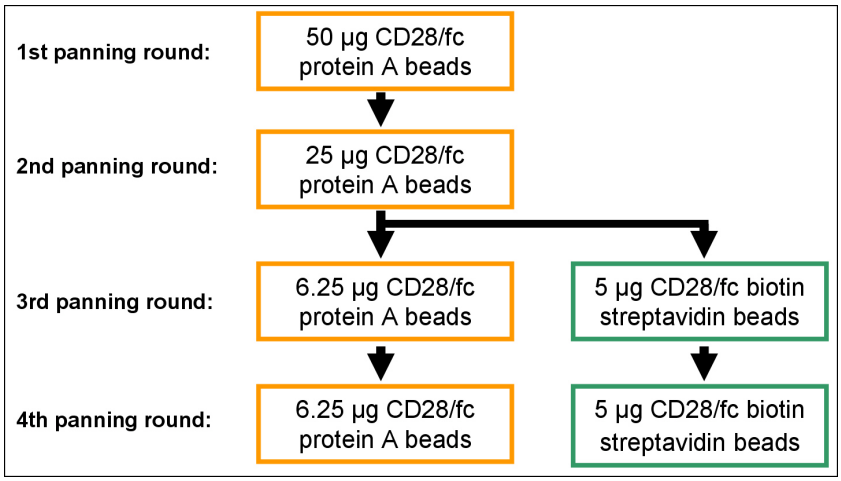


Figure 31: Layout for CD28 panning experiments.
After panning and elution, the remaining phages were amplified and subjected to another round of panning (indicated by the arrows).
After the second round of panning, the amplified phages were screened under two different conditions (two different methods of immobilization) rather than only screening on protein A bead mediated immobilization. The amount of CD28 target shown was the amount used per individual panning experiment.

3.3.3.2 Enrichment of eluted phages

Enrichment was detectable in this panning experiment after the 4th panning round, independent from the immobilization method used – see Table 4. While the output:input ratio finally went up in the experiments with CD28 immobilized, it did not change at all for the background panning experiments (PB #1 and PB #2).

The increase in the output:input ratio for the pannings with CD28 as target is shown in Figure 32. The background panning was left out of this graph.

	4 rounds protein A beads				2 rounds protein A beads, 2 rounds streptavidin beads,	
	P #1	**P #2**	**PB #1**	**PB #2**	**P #1**	**P #2**
1R phage input	2.3×10^{12}	2.3×10^{12}	2.3×10^{12}	2.3×10^{12}		
1R phage output	1.6×10^{6}	2.3×10^{6}	2.0×10^{5}	7.7×10^{6}		
1R output:input	$\mathbf{7.0 \times 10^{-7}}$	$\mathbf{1.0 \times 10^{-6}}$	$\mathbf{8.7 \times 10^{-8}}$	$\mathbf{3.3 \times 10^{-6}}$		
2R phage input	2.0×10^{10}	2.5×10^{10}	2.5×10^{10}	3.2×10^{10}		
2R phage output	1.1×10^{5}	2.8×10^{4}	4.0×10^{4}	2.0×10^{4}		
2R output:input	$\mathbf{5.5 \times 10^{-6}}$	$\mathbf{1.1 \times 10^{-6}}$	$\mathbf{1.6 \times 10^{-6}}$	$\mathbf{6.3 \times 10^{-7}}$		
3R phage input	3.3×10^{10}	2.0×10^{10}	2.7×10^{10}	2.5×10^{10}	3.3×10^{10}	2.0×10^{10}
3R phage output	9.0×10^{4}	1.0×10^{5}	1.7×10^{4}	1.2×10^{3}	8.0×10^{4}	2.0×10^{5}
3R output:input	$\mathbf{2.7 \times 10^{-6}}$	$\mathbf{5.0 \times 10^{-6}}$	$\mathbf{6.3 \times 10^{-7}}$	$\mathbf{4.8 \times 10^{-8}}$	$\mathbf{2.4 \times 10^{-6}}$	$\mathbf{1.0 \times 10^{-5}}$
4R phage input	2.5×10^{10}	3.0×10^{10}	3.2×10^{10}	2.2×10^{10}	2.4×10^{10}	3.0×10^{10}
4R phage output	5.0×10^{6}	3.0×10^{7}	1.1×10^{5}	3.0×10^{3}	6.5×10^{6}	4.0×10^{7}
4R output:input	$\mathbf{2.0 \times 10^{-4}}$	$\mathbf{1.0 \times 10^{-3}}$	$\mathbf{3.4 \times 10^{-6}}$	$\mathbf{1.4 \times 10^{-7}}$	$\mathbf{2.7 \times 10^{-4}}$	$\mathbf{1.3 \times 10^{-3}}$

Table 4: Development of phage titer over four rounds of panning the CPL19YS-2 library on CD28

In the experiments P#1 and P#2 two separate pannings were carried out with the extracellular domain of CD28 as target. PB#1 and PB#2 represent data from pannings on just the protein A beads to check the background enrichment. 1R to 4R indicates the round of panning, input indicates the number of phage incubated with the target protein, while output shows the amount of eluted phage after washing steps.

An increase of the output:input ratio from one round to another usually indicates a successful enrichment of clones presenting ligands with high affinities.

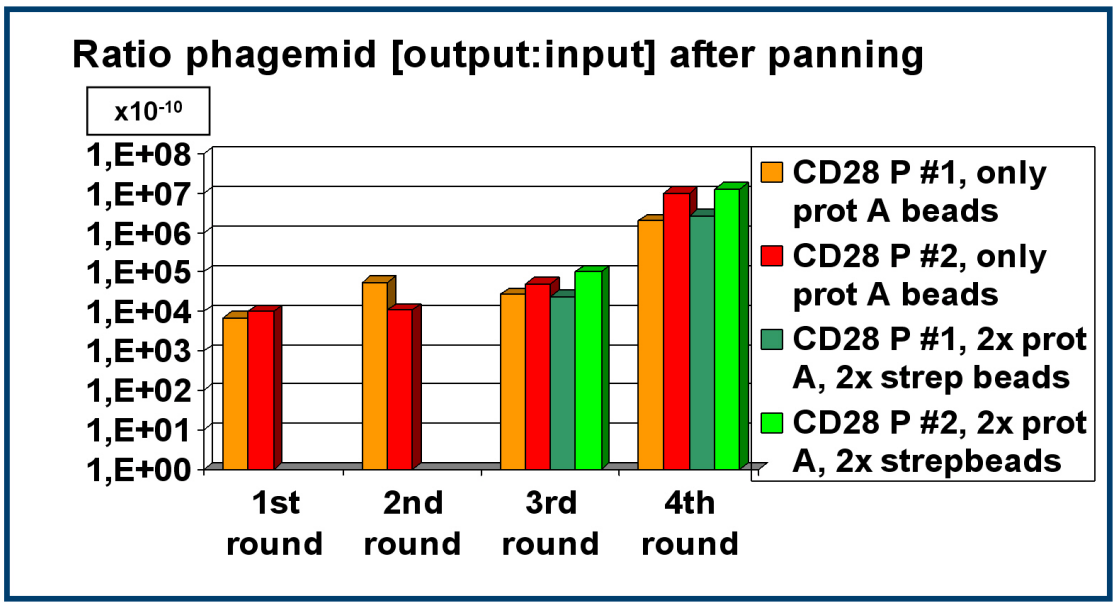


Figure 32: Output:Input ratio for different rounds of panning the CPL19YS-2 library on the CD28 target

3.3.3.3 Analysis of isolated phages

There were two independent panning experiments (P#1 with PB#1 as background control as well as P#2 with PB#2). Single clones were picked and sequenced from each population of isolated clones.

Defective clones were defined to contain mutations that cause either frame shifts or stop codons in the library region. In a few cases, there was no detectable library insert at all. Frequency of these mutations did not show any increase or variation during selection and was in the range of the rates determined at the time the CPL19YS-2 library was created (7.6 % defective clones):

1st panning experiment (4 rounds):

- panning on target: 72/80 clones were correct (10 % defective clones)
- panning on background: 15/16 correct (6.3 % defective clones)

2nd panning experiment (4 rounds):

- panning on target: 68/72 clones were correct (5.5 % defective clones)
- panning on background: 22/24 correct (8.3 % defective clones)

In both panning experiments, enriched sequences were identified (Figure 34b for P#1, and Figure 35b for P#2). A preferential enrichment for certain clones could be observed in round 4 compared to round 3.

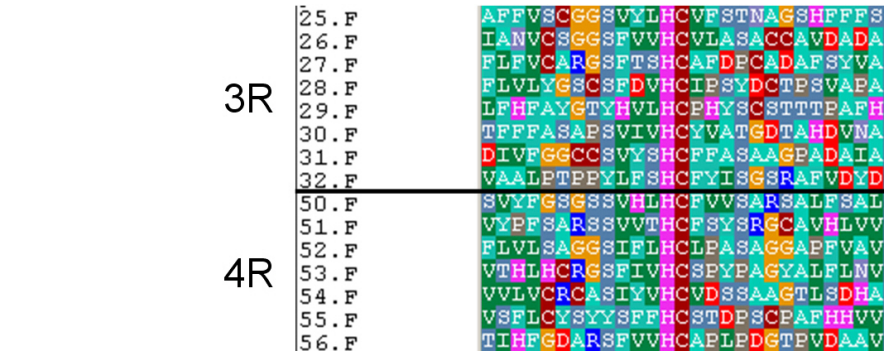
It is worth noting that the change of the immobilization procedure seems to have caused a change in the preferred enrichment conditions — while all 3rd rounds of panning were run with the same phagemid population obtained from the 2nd round of panning, different clones were subsequently enriched. This can be concluded from comparing the clones isolated already after three panning rounds, i.e. one must summarized that some clones had been specifically enriched for affinity to the protein A matrix material in combination with the immobilized antibody.

The sequencing of the clones obtained from the "background" panning (panning CPL19YS-2 against protein A beads, without any target captured) showed no specific enrichment of single clones or similar consensus sequences (Figure 34a for PB#1, and Figure 35a for PB#2), this result fits the lack of increase in the output:input ratio of phage during the consecutive panning rounds (Figure 32).

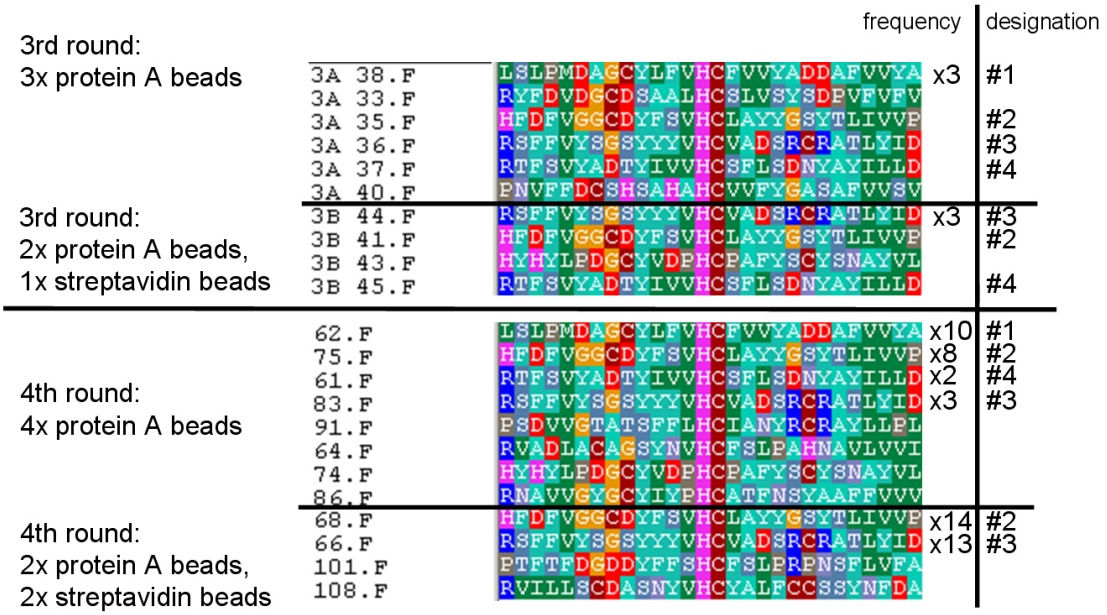
A comparison of the enriched sequences does not show obvious similarities or consensus sequences of more than one or two amino acids and therefore no definite motifs (see Figure 33). The exception is a rather small common sequence of VYAD that comes up in the clones #1 and #4. It seems worth noting that neither clone #1 or clone #4 were enriched in the pannings where protein A as immobilization matrix was replaced by streptavidin, as mentioned above.

Looking at the potential for the formation of disulfide bridges, 6 of 7 enriched sequences possess an even number of cysteines (1×single cysteine, 5×double cysteine, 1×four cysteines). As mentioned by McConnell et al. (1996) and Kay et al. (1993), due to lack of further data it can not be concluded if this is due to preferential enrichment *for* clones that can form an intramolecular disulfide loop ("rigidifying" the peptide) or just a selection *against* clones which possess an unpaired cysteine, potentially forming inappropriate disulfide linkages between the displayed peptide and other cysteine residues. The clones enriched in the evaluation panning on anti-p53 Ab-6 (section 3.3.2) did not show this preference, as did the original CPL19YS-2 library, indicating that the preference for paired cysteines are specific to the panning on the CD28 target.

Figure 33: Overview of sequences enriched in two independent experiments of panning the CPL19YS-2 library on the extracellular domain of CD28 (after four rounds of panning each)
Numbers #1–#7 refer to the numbering of clones enriched after the final round in the two separate panning experiments. Clones #1–#4 were enriched in panning P#1, the clones #5–#7 were isolated in panning P#2. The frequency gives the number of clones isolated in the fourth panning round. See Figures 34 and 35 for detailed panning results.

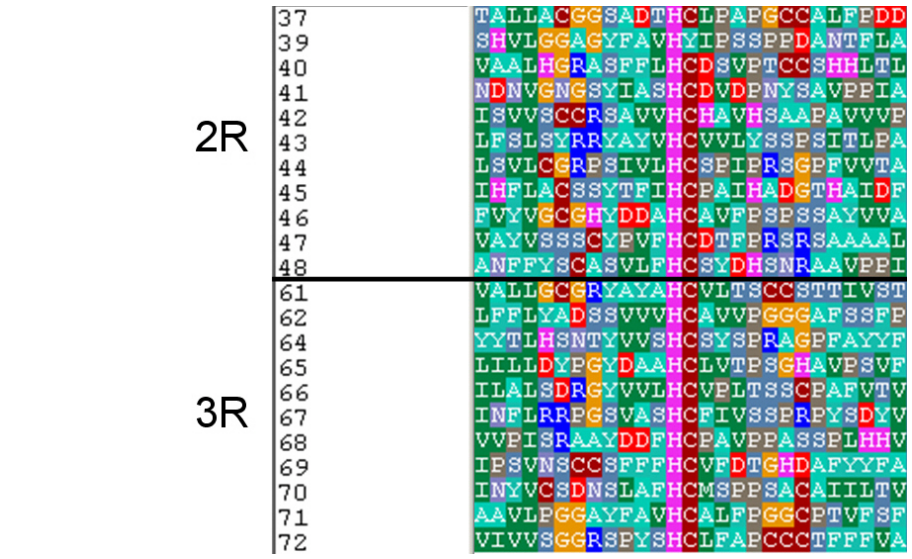


(a)

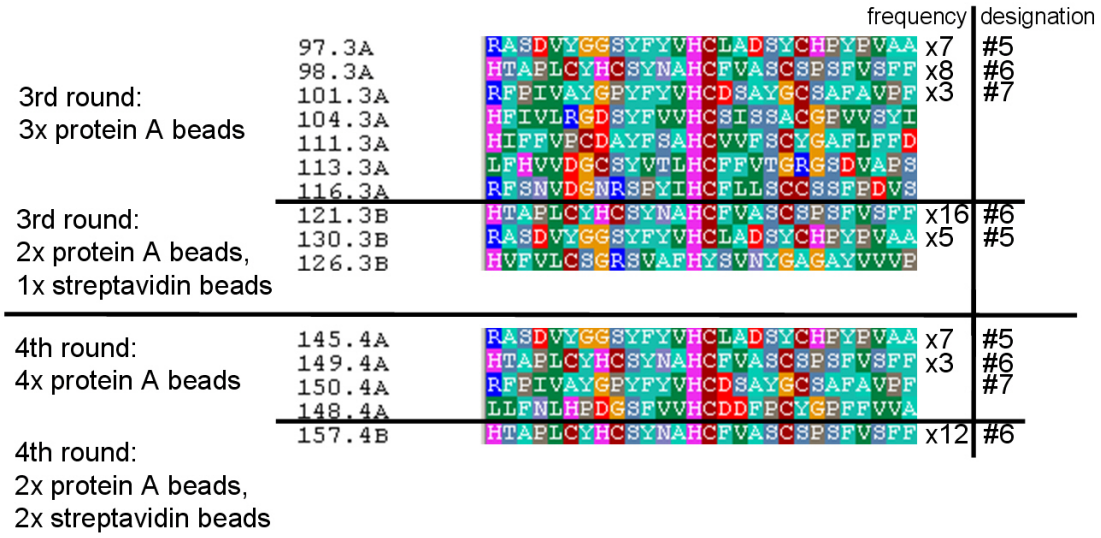


(b)

Figure 34: First panning experiment on CD28:(a) Sequences of single clones isolated after 3/4 rounds of panning on protein A beads (background for CD28 panning), panning PB#1. 3R indicates 3rd panning round, 4R means 4th panning round. No single clone or a consensus sequence of different clones could be detected.
(b)Sequences of single clones isolated after 3/4 rounds of panning the CPL19YS-2 library on CD28, panning P#1. Numbers #1–#4 indicate clones that appear more than once and thus are considered to be enriched rather than just being picked coincidentally. Sequences isolated more than once are shown with their frequencies of isolation.



(a)



(b)

Figure 35: Second panning experiment on CD28:(a) Sequences of single clones isolated after 2/3 rounds of panning on protein A beads (background for CD28 panning), panning PB#2. 2R indicates 2nd panning round, 3R means 3th panning round. No single clone or a consensus sequence of different clones could be detected.
(b)Sequences of single clones isolated after 3/4 rounds of panning the CPL19YS-2 library on CD28, panning P#2. Numbers #1–#4 indicate clones that appear more than once and thus are considered to be enriched rather than just being picked coincidentally. Sequences isolated more than once are shown with their frequencies of isolation.

3.3.3.4 Recombination experiment

After the enrichment and identification of several clones from panning the CPL19YS-2 library on the CD28 target, the cassette structure of the library should be used to examine closer which part of the sequences was contributing most to the enrichment in the panning and therefore probably was also contributing most to the peptide's affinity to the CD28 target.

By "backcrossing" the enriched clones with the original library and panning the resulting new library on CD28 again, clones should be preferentially enriched that contain the sequence parts contributing most to affinity.

In order to achieve this, either the first two or the latter two of the four cassettes from the enriched clones from the P #1 panning on CD28 (clones #1 62, #2 75, #3 83, #4 61 from Figure 33) were replaced with the highly varied cassettes from the original library. The principle is explained in Figure 36.

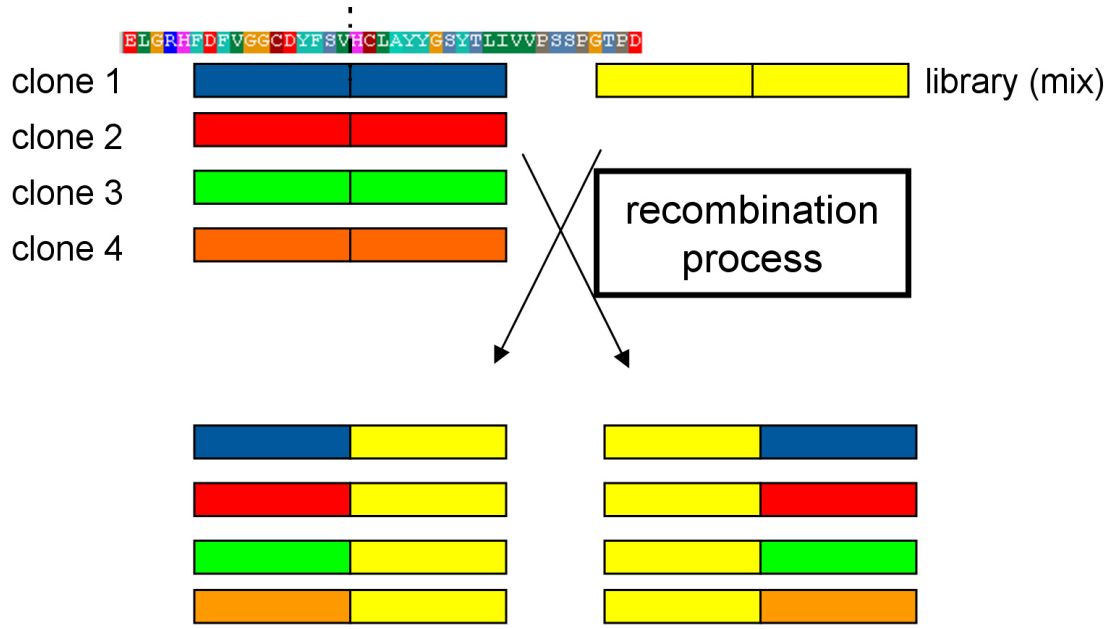


Figure 36: Principle used for recombination of four selected DNA sequences with the sequence variety in the whole CPL19YS-2 library.
The colors *blue, red, green* and *orange* represent specific DNA sequences from specifically isolated clones #1–#4. The *yellow* bar represents the whole complexity of variable DNA sequences existing in the CPL19YS-2 library. A recombination between each left sequence parts of one pool with the right sequence parts of the other pool results in new DNA sequences which possess one half consisting of just four different sequence parts, with the other half consisting of various different sequence parts from the library.

In order to carry out this recombination procedure (see Figure 37 for an overview), plasmid DNA was isolated from Top10 F'λ carrying the CPL19YS-2 library in plasmid format as well as from Top10 F'λ cultures carrying the clones #1 62, #2 75, #3 83 and #4 61. DNA concentration of all plasmid preparations was determined and the single clone plasmid preparations were pooled to get a 1:1:1:1 molar mix ratio of the clones #1–#4.

Plasmid DNA from the library mix as well as the plasmid DNA from the four clones was digested with two different pairs of restriction enzymes, either *Bsr*DI+*Bgl*I or *Bsr*DI+*Dra*III. Digestion with the first pair of enzymes resulted in two DNA fragments (1514 bp and 3188 bp size), digestion with the second pair also gave two DNA fragments (1855 bp and 2847 bp size).

From the library plasmid digestions, the larger fragments (3188 bp respective 2847 bp) were purified by agar gel electrophoresis and following extraction from the gel. From the clone mix plasmid digestions, the smaller DNA fragments (1515 bp respective 1855 bp) were separated and purified in the same way.

The purified fragments were then ligated in the following combination:

- The large library fragment and the small clone mixture fragment from the *Bsr*DI+*Bgl*I digestion (3188 bp + 1515 bp)
- and the large library fragment and the small clone mixture fragment from the *Bsr*DI+*Dra*III digestion (2847 bp + 1855 bp)

After a cleanup procedure, the DNA was transformed into Top 10 F'λ cells by electroporation. The final number of clones obtained in this way was

- **2×10^7 for the *Bsr*DI+*Bgl*I recombination** (right sequence part varied, left part from just the four clones — further designation for the resulting library: CPL19YS-2-CD28B), and
- **7×10^6 for the *Bsr*DI+*Dra*III recombination** (left sequence part varied, right part from just the four clones — further designation for the resulting library: CPL19YS-2-CD28D).

Both populations were amplified and packaged into phagemid format in order to get two new libraries that could be used for panning experiments.

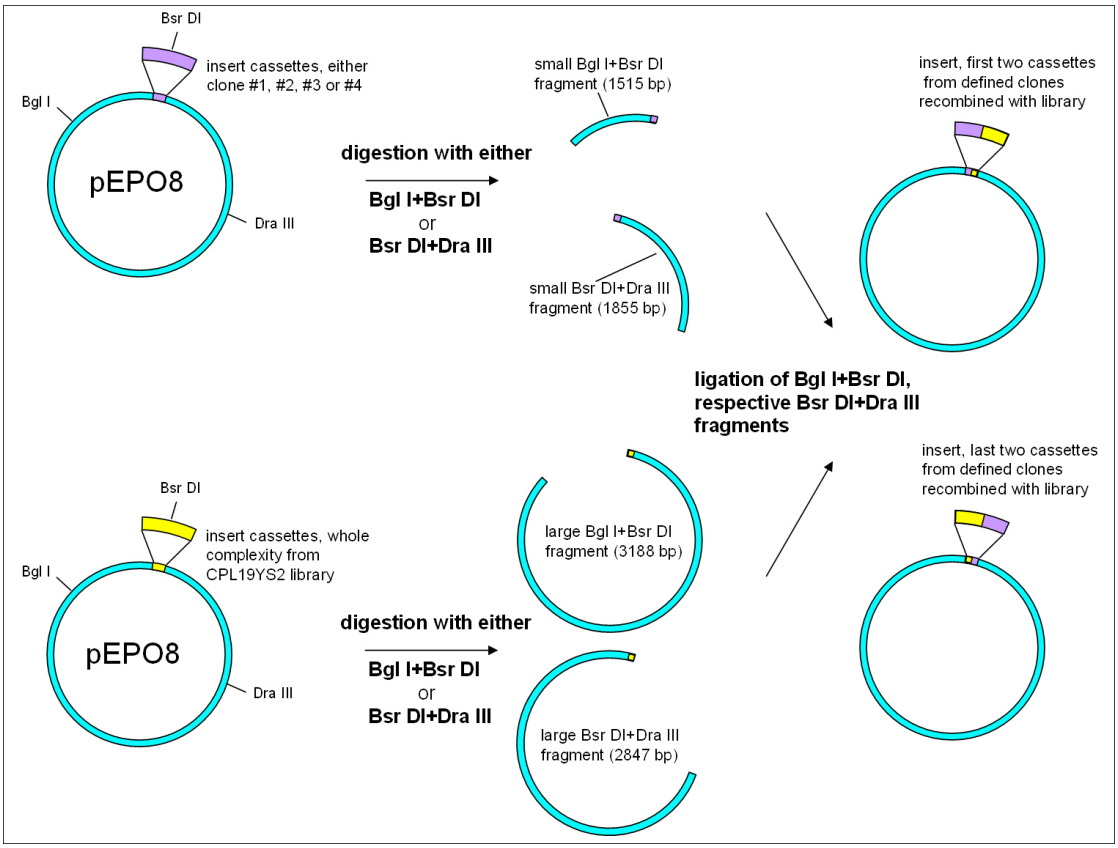


Figure 37: Digestion and ligation scheme used for recombination of library cassettes from distinct clones with variable cassettes from the whole CPL19YS-2 library.
The *violet* color represents the DNA cassettes from the four isolated clones #1–#4, the *yellow* cassettes represent the variable DNA sequences present in the entire CPL19YS-2 library. Unique recognition sites for restriction enzymes are indicated.
Phagemids, either pooled from plasmid DNA preparations of four previously isolated clones (upper part) or from plasmid DNA preparations of the CPL19YS-2 library (lower part) were digested using two different sets of restriction enzymes. Either the small or the large resulting fragments were isolated and then the respective fragments from both experiments were ligated to create recombined phagemids that contain a noted combination of library cassettes.

Panning conditions for panning the recombined libraries on CD28 The panning conditions used to pan the new libraries [designated CPL19YS-2-CD28B and CPL19YS-2-CD28D] on the CD28 target were the same as in section 3.3.3.1.

A background control panning was carried out to detect unspecific binding, this was the panning PB#2 from section 3.3.3.1.

Enrichment of eluted phages Enrichment was detectable in this panning experiment already the 3rd panning round, independent from the immobilization method used — see Table 5. Compared to the other two pannings on CD28 using the original CPL19YS-2 library (see section 3.3.3.2), this enrichment occurred earlier. The final output:input ratio was about the same in the 3rd and 4th round of panning, which implies that the most affine ligands had already been enriched.

The increase in the output:input ratio for the pannings with CD28 as target is shown in Figure 38.

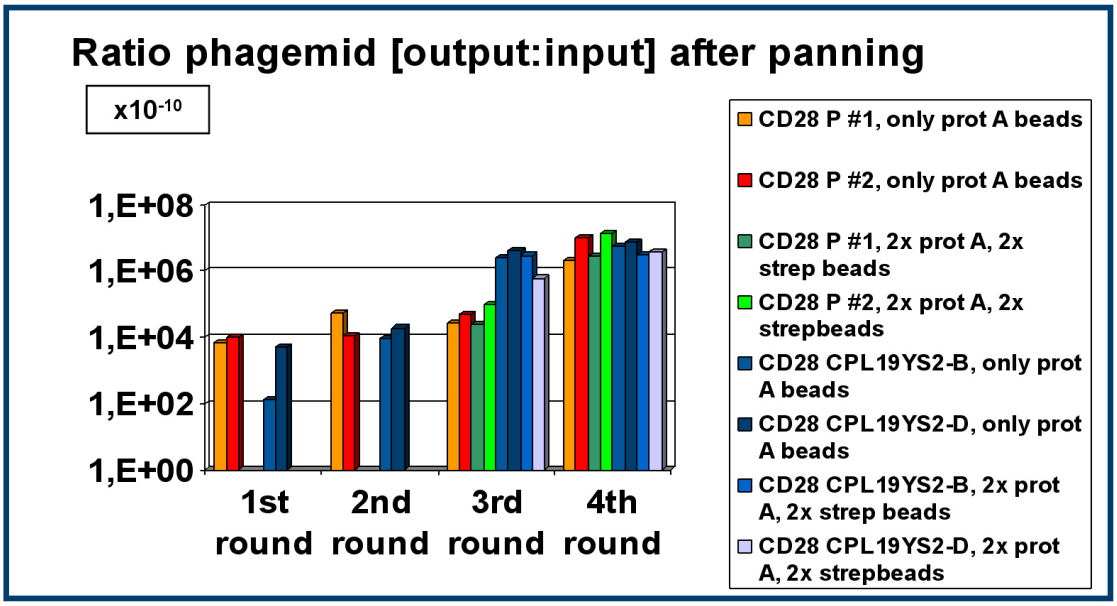


Figure 38: Output:Input ratio for different rounds of panning either the CPL19YS-2 library (P #1 and P #2) or the recombined CPL19YS-2-B/-D libraries on the CD28 target
The ratios for the pannings of the CPL19YS-2 library on CD28 (green and red/orange bars) do all show more or less the same development, while the panning experiments of the recombined CPL19YS-2-B/-D libraries show an earlier increase in the output:input ratio (already in the 3rd panning round, compared to the 4th round for the CPL19YS-2 pannings).

	4 rounds protein A beads		**2 rounds protein A beads, 2 rounds streptavidin beads,**	
	CD28-B lib	**CD28-D lib**	**CD28-B lib**	**CD28-D lib**
1R phage input	3.0×10^{12}	2.0×10^{11}		
1R phage output	4.0×10^{4}	1.0×10^{5}		
1R output:input	$\mathbf{1.3 \times 10^{-8}}$	$\mathbf{5.0 \times 10^{-7}}$		
2R phage input	4.0×10^{10}	3.5×10^{10}		
2R phage output	3.5×10^{4}	6.5×10^{4}		
2R output:input	$\mathbf{8.8 \times 10^{-7}}$	$\mathbf{1.9 \times 10^{-6}}$		
3R phage input	2.5×10^{10}	3.0×10^{10}	2.5×10^{10}	3.0×10^{10}
3R phage output	6.5×10^{6}	1.2×10^{7}	7.2×10^{6}	1.8×10^{6}
3R output:input	$\mathbf{2.6 \times 10^{-4}}$	$\mathbf{4.0 \times 10^{-4}}$	$\mathbf{2.9 \times 10^{-4}}$	$\mathbf{6.0 \times 10^{-5}}$
4R phage input	2.4×10^{10}	2.8×10^{10}	3.0×10^{10}	2.8×10^{10}
4R phage output	1.3×10^{7}	2.0×10^{7}	9.0×10^{6}	1.0×10^{7}
4R output:input	$\mathbf{5.4 \times 10^{-4}}$	$\mathbf{7.1 \times 10^{-4}}$	$\mathbf{3.0 \times 10^{-4}}$	$\mathbf{3.6 \times 10^{-4}}$

Table 5: Development of phage titer over four rounds on panning the CPL19YS-2-B/-D libraries on CD28

CD28-B lib and CD28-D lib indicate the library that was the source for the phages used in the first round of panning (CPL19YS-2-B or CPL19YS-2-D). 1R to 4R indicates the round of panning, input indicates the number of phage incubated with the target protein (the extracellular domain of CD28), while output shows the amount of eluted phage after washing steps. An increase of the output:input ratio from one round to another usually indicates a successful enrichment of clones presenting ligands with high affinities. As explained in section 3.3.3.1, the immobilization method for retaining the target protein was varied, indicated by the top line of the table.

Analysis of isolated phages In this panning experiment, single clones were picked and sequenced for their library DNA sequence starting after the 2nd round of panning.

The frequencies of mutation (frame shift/stop codon mutations in the library region or no detectable library insert at all) did not vary over the course of the panning experiments and were even lower (144 clones sequenced, 4.1 % of them were defective) than the rates determined throughout the creation of the CPL19YS-2 library (7.6 %). Due to the relatively low number of experiments, it can not be concluded if this low rate is due to a recombination of the original CPL19YS-2 library with clones that have been completely verified for a correct sequence. After the 2nd panning round (see Figure 39), most of the original kept parts of the recombined clones #1–#4 could still be found in the sequences of the clones analyzed: From the CPL19YS-2-B library, all four different left sequence parts could be detected, while for the CPL19YS-2-D library only two different right sequence parts were found. Four specific clones were already found multiple times, indicating enrichment of certain phagemids. A selection and enrichment had obviously already happened in the two panning rounds, an enrichment that "early" in the panning experiments had not been observed in the two panning experiments of the CPL19YS-2 library on CD28.

As it was also in the case of the panning of the CPL19YS-2 library on CD28 (section 3.3.3.3), the modification of the immobilization procedure seems to have caused a change in the preferred enrichment conditions — while the input for the two different pannings in the 3rd panning round was the same phagemid population obtained from the 2nd round of panning, different clones were subsequently enriched. This effect is shown in Figure 40 for the panning with the CPL19YS-2-B sublibrary and in Figure 41 for the panning with the CPL19YS-2-D sublibrary on CD28.

After four rounds of panning, both sublibrary pannings yielded not just three or four single specific clones (as with the pannings of the original CPL19YS-2 library on CD28, compare section 3.3.3.3), but rather a multitude of different clones (with some of these occurring more than once, see Figures 40 and 41).

The clones enriched were sorted in groups with different parts of the original recombined sequences. The designations chosen for these groups were 1B-4B for the four different original left sequence parts (for the recombined clones #1–#4) in the CPL19YS-2-B library and 1D–4D for the four different original right sequence parts in the CPL19YS-2-D library.

This result has to be analyzed under two different aspects — for one thing, this experiment has been conducted to examine the contribution of the pre-selected sequences to the binding to the CD28 target. If one half of a preselected sequence appears more often after four rounds of enrichment than the other half, it is highly probable that it contributes more to affinity than the other half which was less frequently enriched in the parallel experiment.

Another point is the screening for synergistic motifs by recombining the preselected sequences with the original library — while one half of the pre-selected sequence gets preferentially enriched over the other, the variable sequence part from the library (Figure 36) will be screened for motifs that contribute further affinity to the pre-selected sequence halves, "enhancing" their binding properties.

If the "groups" of enriched clones shared the affinity/enrichment relevant motifs, that could indicate that the actual conserved "half" of the recombined clones that is defining the group designation was contributing most to the affinity — which was one point in designing that experiment (see section 3.3.3.4).

Looking at the *groups* of sequences that were enriched, not at single clones, it is obvious that *different* groups were preferentially enriched: In the 4th panning round of the CPL19YS-2-B library, only 1 out of 24 clones belonged to group 1B, the other 23 clones were all from group 4B. This means that the left part of the original clone #4 was predominantly enriched over the other groups.

In the panning of the CPL19YS-2-D library, clones from the groups 1D, 2D and 3D were enriched finally (groups 1D and 3D for the panning on 4×protein A beads, groups 2D and 3D for panning on 2×protein A beads and 2×streptavidin beads). No clone belonging to group 4D could be found among the analyzed clones.

Checking all isolated clones from the 4th panning round, 43 of 44 sequences had an even number of 2 cysteines, only one single clone (#228 in the panning of the CPL19YS-2-B library) was found that just possessed the central fixed cysteine, without any more cysteine in the flanking left or right sequence parts. As already discussed for the first two panning experiments on CD28, the selection for paired cysteines indicates advantageous properties for those clones (like a "rigidified" peptide structure or the "presentation" of an amino acid motif between two cysteines), although this can not be confirmed without direct comparison experiments or structural data.

A comparison of the enriched sequences does not show obvious similarities or consensus sequences of more than one or two amino acids and therefore no *definite* motifs.

However, as already noted in section 3.3.3.3 from the pannings of the original CPL19YS-2 library, the VYAD motif is part of the sequence in the two most enriched groups in the pannings (group 4B, Figure 40 and group 1D in Figure 41). As in the first panning experiments on CD28, in the panning of the sublibrary CPL19YS-2-D the group of clones containing the VYAD motif was selected against in the 3rd/4th round of panning after switching from protein A beads to streptavidin beads. In the panning of the CPL19YS-2-B library on CD28, the group of clones containing VYAD was still present after the switch of the immobilization matrix. This could mean either that VYAD is not specific for the protein A matrix in combination with the target, or that the first two rounds

of enrichment on protein A beads already enriched this group of clones enough to stay dominant after the change of the target immobilization method.

In the group of clones enriched from the CPL19YS-2-B library a weak consensus sequence can be assumed after four rounds of panning (Figure 40). This consensus sequence is S(L/F)(L/F)(H/N/P)GGCAA, but has not appeared at any other position in the enrichment. The loop length of the (potential) disulfide loop is seven /six amino acids, enclosing mostly small amino acids .

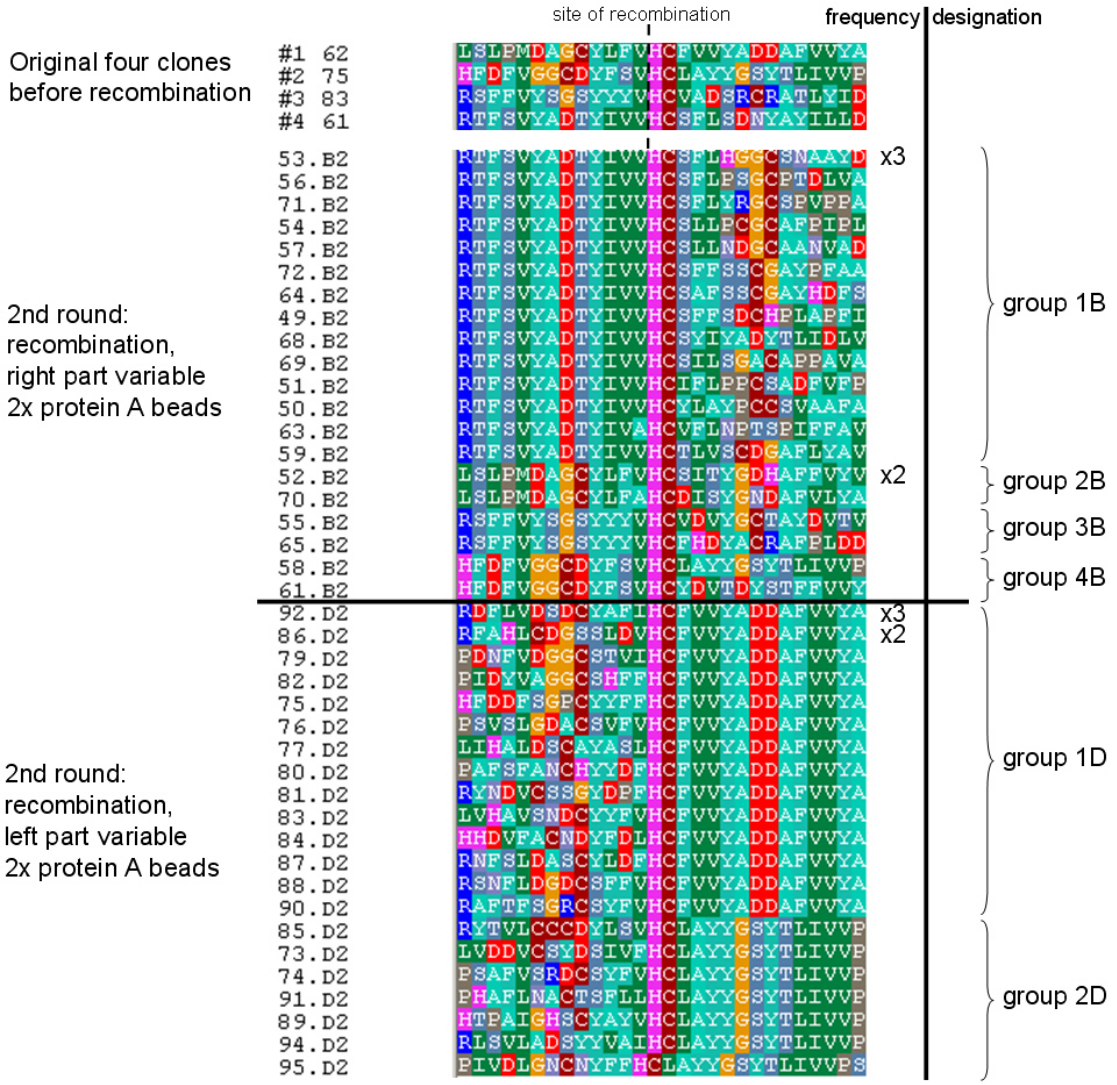


Figure 39: Sequences of single clones isolated after 2 rounds of panning the recombined libraries CPL19YS-2-B and CPL19YS-2-D on CD28
"grouping" was done based on from which original clone the fixed sequence part was (B for clones with the left parts kept, from the CPL19YS-2-B library, D for clones with the right parts kept, from the CPL19YS-2-D library). Sequences isolated more than once are shown with their frequencies of isolation.

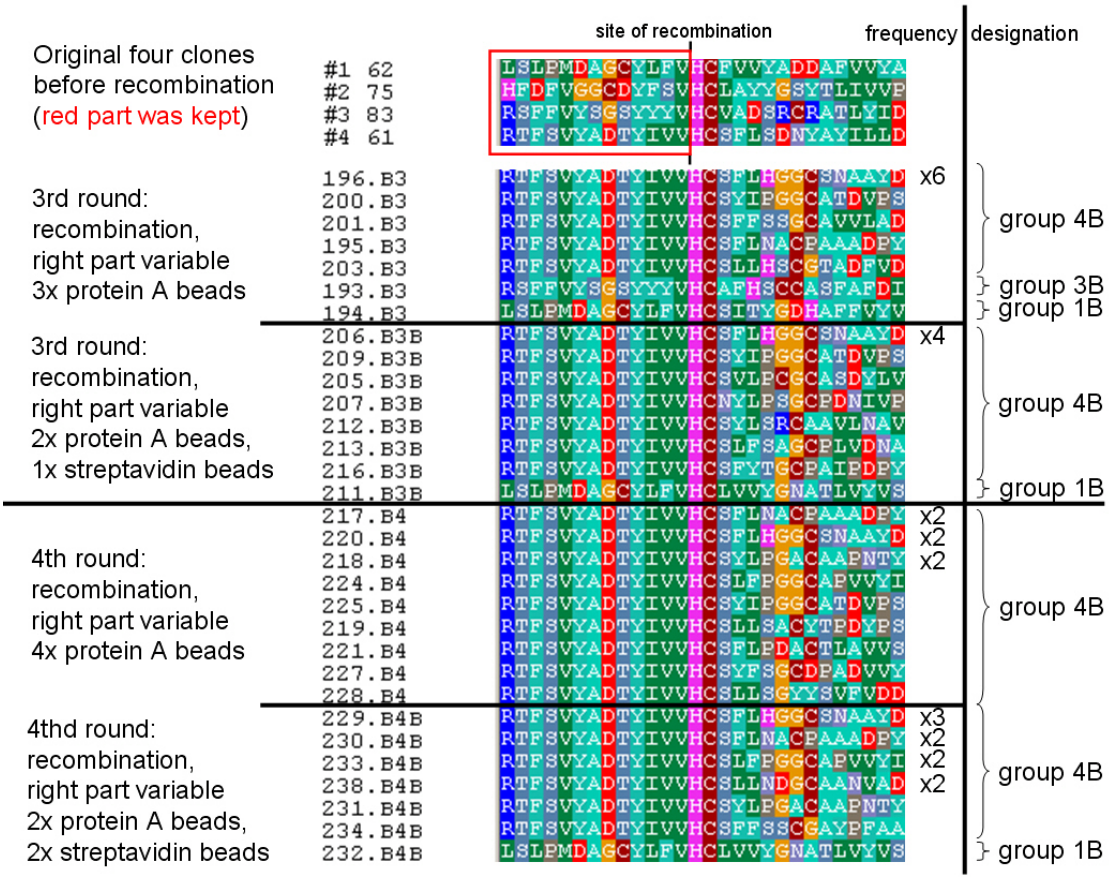


Figure 40: Sequences of single clones isolated after 3/4 rounds of panning the recombined library CPL19YS-2-B on CD28

"grouping" was done based on from which original clone the fixed sequence part was, B indicates origin from the CPL19YS-2-B library. Sequences isolated more than once are shown with their frequencies of isolation. A weak consensus sequence can be assumed in the sequences isolated after four rounds of panning, especially in the last seven sequences. It starts after the central fixed HC position and is S(L/F)(L/F)(H/N/P)GGCAA, but has not appeared at any other position in the enrichment. The loop length of the (potential) disulfide loop is mostly seven amino acids, enclosing mostly small amino acids.

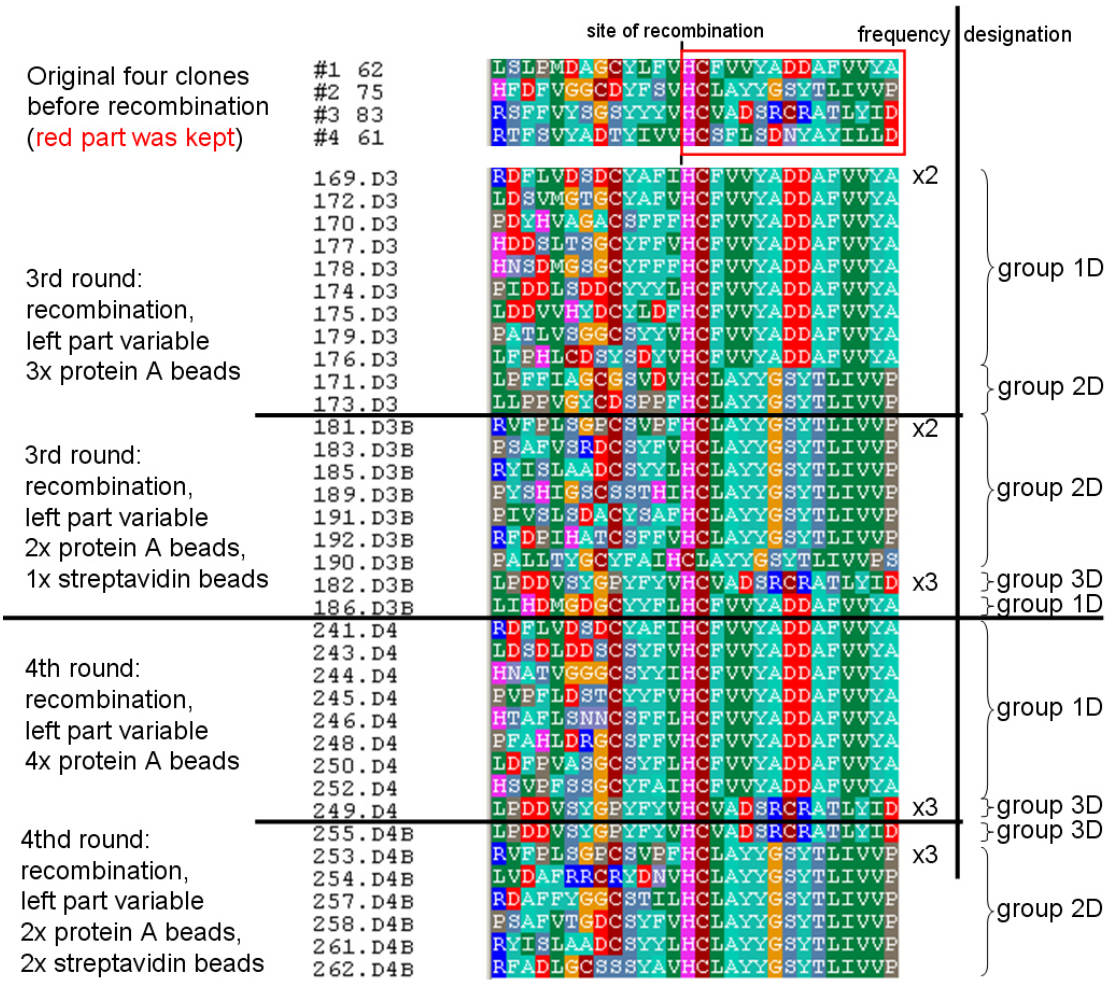


Figure 41: Sequences of single clones isolated after 3/4 rounds of panning the recombined library CPL19YS-2-D on CD28

"grouping" was done based on from which original clone the fixed sequence part was, D indicates origin from the CPL19YS-2-D library. Sequences isolated more than once are shown with their frequencies of isolation.

3.3.3.5 Conclusions from panning on CD28

The two independent panning experiments gave comparable results, although the isolated clones did not show obvious similarities. Both independent pannings took about the same number of panning rounds to enrich the clones finally perceived as dominant; the fact that the sequences are not looking similar can not be explained finally — it remains open if this is either due to an original library complexity that was too high to allow reproducible enrichments: The original fraction of phages used for the first round of screening in phage display needs to contain at least 10^4 clones of each presented peptide in order to reproduce enrichments (from personal communication with John Collins).

The complexity used here, 4.2×10^8 definitely approached that limit, considering the amounts of phage used in the initial panning rounds (2.3×10^{12}). This result further emphasized the considerations regarding the desired complexity of a primary library, as discussed in secton 3.1.1.

None of the panning experiments on CD28 yielded any detectable enrichment of target-unspecific motifs. While this eliminated the problem of having to determine which enriched candidates were matrix- and which were target-specific, the question arises why no clones from the library were showing affinities to the immobilization material that was common in some experiments (polystyrene *Dynabeads* and the recombinant protein A that was covalently coupled to the beads, in some pannings used in four consecutive rounds). A minor point could be that the majority of motifs enriched on protein A (according to Menendez and Scott, 2005) contained tryptophan, which is absent in the library.

The two different immobilization methods used seemed to have an impact on the preference on which clones were finally enriched. While this change of the immobilization method for the target used for panning (section 3.3.3) had originally been implemented to reduce the enrichment of target-unspecific phages, it also introduced a slight change in the structure of the target itself, since the recombinant fusion protein used in the first two panning rounds contained the amino acid residues 1–152 of CD28, and the biotinylated fusion protein that was used partially in the rounds 3–4 contained only amino acid residues 1–134.

This slight difference in the presented target was necessary due to the commercial availability of the target, however, it also showed that the target structure *had* a clear effect on the preference for certain clones. Were structural information about the isolated clones available (see also section 3.4 for prediction algorithms inferring secondary structure likelihood), which is not, one could try to relate the change in preference to the change of target structure.

Right now the final proof remains to be given that the isolated clones actually ex-

hibit affinity to the CD28 target, even without the surrounding structuring matrix the phage surface might provide. The unsuccessful experiments to analyze target-peptide interactions are presented in section 3.5.

3.3.3.6 Effect of recombination on the panning process

Due to time and material constraints, the panning on CD28 using the recombined clones was only performed once, so the reproducibility of the outcome can not be evaluated. The fact, however, that a strong enrichment was observed, implies at least that the two recombined libraries are usable for screening.

Even though accurate measurement values were not possible (e.g. from Biacore experiments, see section 3.5), the observed results indicate that the recombination of clones isolated from a panning and "backcrossed" with the original library indeed was advantageous for the affinity properties of clones in the library, since the output:input ratio of phages went up earlier in the panning cycle using the recombined libraries than it did for the original CPL19YS-2.

This does however not give information if the recombined clones are actually *more* affine, or if the new sublibraries just have a greater portion of clones that express affinity to the target (due to the motifs contributed by the clones enriched on CD28 before).

3.3.4 Evaluation pannings on other targets

In addition to the pannings on the anti-p53 Ab-6 antibody and CD28, there were two more panning experiments – on streptavidin coated beads *Dynabeads M-280 Streptavidin, Invitrogen, California*) to select for clones specific to the streptavidin surface, and on the M3.5 protein (a part of the M3 protein of group A streptococci, Dinkla et al., 2007, 2009), which was expressed as a GST (glutathione S-transferase) fusion protein and immobilized on glutathione sepharose beads (Glutathione Sepharose 4 Fast Flow, *GE Healthcare, United Kingdom*).

The panning of the CPL19YS-2 library on streptavidin beads was performed similarly to the pannings on CD28 without anything immobilized on the beads, to get comparable results and determine potential streptavidin specific sequence motifs in the library. The panning on the GST-M3.5 fusion protein was performed on amounts of 10–80 µg of target in PBS, following the conditions stated by Dinkla et al. (2007).

The library was panned twice on both targets for four rounds, but no enrichment was detected, neither an increase in the phage output:input ratio, nor any enrichment of specific single clones. After four rounds, mainly clones had been enriched that contained deletions in the library region, leading to no expression and presentation of fusion protein at all.

The enrichment of clones encoding no (presentable) fusion protein at all is usually a strong indication for no target-specific enrichment at all, since phagemid infected bacteria without an encoded fusion protein have a slight growth advantage. The absence of any increase in the phage output:input ratio after each panning round further supports that conclusion, suggesting that the CPL19YS-2 library does not contain clones with high affinities to both targets.

3.3.5 Evaluation panning on a polyclonal serum recognizing a specific linear *Pseudomonas* peptide antigen

Since the panning of the CPL19YS-2 library on the antibody anti-p53 Ab-6 (section 3.3.2) had only yielded a single clone with an expected motif enriched, another evaluation panning on antibodies was conducted. This work was carried out by Juliane Lindner and Jonas Kügler (Helmholtz Centre for Infection Research). The results are, however, presented here, as they provide valuable data for the evaluation of the library with respect to diversity and quality.

The panning target is a polyclonal mouse serum against a peptide from *Pseudomonas aeruginosa* with the sequence CVAPTADEHFTTLFQATNPS. It was expected that related, perhaps overlapping motifs would be enriched during panning the CPL19YS-2 library.

3.3.5.1 Panning conditions

The panning was performed as described in chapter 2.2.5.1, except for the washing conditions, which were changed to ten times washing with PBS-T. The amount of target immobilized for each panning was 50 µL of serum, immobilized on 50 µL of protein A beads. Four rounds of panning were performed, the clones enriched after the fourth round were analyzed by sequencing.

3.3.5.2 Analysis of isolated phages

After four rounds of panning, 32 clones were analyzed per sequencing. Of these, 2 were "defective" in that they had a stop codon mutation in the library sequence, leading to no expression of the peptide-pIII fusion (6.25 % defective clones, in line with the other pannings).

In contrast to the panning on the antibody anti-p53 Ab-6 (section 3.3.2), a large set of different clones was enriched in this experiment, and all of them sharing a region of sequence homology, allowing the proposal of a consensus sequence for the region probably contributing the most to the affinity to the polyclonal antibodies (Figure 43a).

It can also be noted that a single clone, # 19 (Figure 43b), was enriched that contains a motif of 10 aa length which is almost identical to a 9 aa stretch of the *Pseudomonas* antigen with the exception of the red A position and the insertion of a D at the fourth position. This clone's sequence is the one that is closest to the *Pseudomonas* peptide against which the polyclonal antibodies were generated.

The only position where no amino acid in the consensus sequence corresponds with a respective amino acid in the *Pseudomonas* peptide is the glutamic acid in the marked blue region, where only alanine, aspartic acid or leucine were found in the enriched clones at this position. As no glutamic acid was intended to be encoded in the variable part of the library, it would not be available for enrichment.

However, considering that aspartic acid has a similar hydrophilicity a positive charge under the pH of the panning buffer, aspartic acid can be considered an acceptable substitution for glutamic acid at this position.

Figure 42: Sequences of single clones isolated from the CPL19YS-2 library after 4 rounds of panning on a polyclonal antibody serum against the *Pseudomonas* peptide CVAPTADEHFTTLFQATNPS
The amino acid sequences have been aligned with the *Pseudomonas* peptide sequence to allow the recognition of a consensus sequence. Clones isolated more than once are shown with their frequencies of isolation.

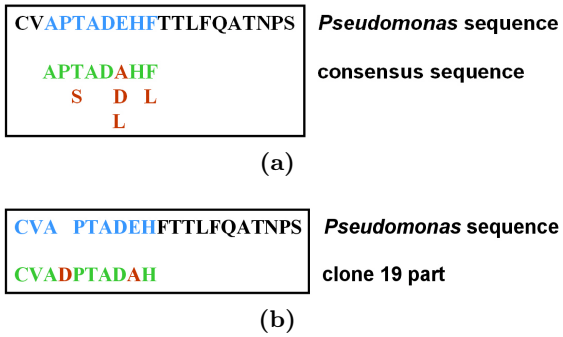


Figure 43: Motifs obtained from panning the CPL19YS-2 library on a polyclonal antibody serum against a *Pseudomonas* peptide (a) Consensus sequence from 13 different clones compared to the original *Pseudomonas* antigen (b) The clone 19 has a motif of 10 aa length that is almost identical to a 9 aa stretch of the *Pseudomonas* antigen with the exception of the red A position and the insertion of a D at the fourth position.

3.3.5.3 Conclusions from panning on a polyclonal serum recognizing a specific *Pseudomonas* peptide

The results from this panning experiment complement the previous validation screening of the CPL19YS-2 library on the anti-p53 Ab-6 antibody. They further confirm that panning & selection is working. This is an independent validation of the quality of the new library. In comparison to that previous panning, a variety of different clones was enriched which all shared a long (7/8 aa) easily identifiable consensus sequence.

This enrichment of clones with a similar or identical binding sequence (consensus motif) with different flanking sequences is expected from a highly diverse library. This supports the final (assumed) conclusions drawn after the panning on the anti-p53 Ab-6 antibody, stating that the lack of diversity in the obtained clones from that experiment was rather due to too stringent washing conditions than to a lack of diversity in the library. Previous uncertainty of library quality based on earlier experiments are thus considered invalidated.

The statistical probability that a clone, randomly picked from the library, will contain the enriched consensus sequence (Figure 43a) can be calculated estimated:

Averaging ten different possible amino acids per variable position, the positions 1(A), 2(P), 4(A), 5(D) and 7(H) would each contribute a 10 % probability (only one defined amino acid), while positions 3(T/S) and 8(F/L) would each give a 20 % probability (two defined amino acids from ten possible). These probabilities, multiplied by five as the number of overall five possible positions that might hold the consensus due to limitations in library composition, result in a probability of $\frac{5}{10^5 \times 5^2} = 2 \times 10^{-6}$. Multiplied with the (assumed) library complexity of 1.8×10^8 (section 3.2.2.3), one could expect about 360 different clones in the library expressing the consensus sequence. 12 clones have been enriched dominantly in the panning when 32 clones have been analyzed after four panning rounds, there may be more different clones present in lower concentrations. Using an additional correction factor of 3 due to the (statistically low) sampling number of just 32 clones sequenced (Collins, 1997), one comes to the final result that there would be about 36 ($=12 \times 3$) clones enriched in the experiment, which means that ~10 % of 360 clones with the expected motifs got enriched in this screening.

It is noted that the epitopes isolated here from the CPL19YS-2 library are linear motifs; 11 out of 12 enriched clones present that motif outside of a (postulated) disulfide stabilized loop. While disulfide bridges confer stability in general, and in fact 10 out of 12 clones have an even number of cysteines, one would expect linear motifs to be enriched when panning against polyclonal antibodies that were selected against a *linear* peptide.

Thus, this test did not validate the efficiency of the library for isolating complex motifs with secondary structures, but the findings are consistent with the expected library complexity, stability of the library and a reasonable efficiency of display.

3.4 Computer analysis of secondary structure elements in peptides from the library

It is generally expected that an enrichment of binding motifs is enhanced by the ligand forming a secondary structure. Being able to characterize secondary structures formed would therefore help to evaluate binding motifs. Since accurate predictions of the secondary structure are not possible based solely on the amino acid primary structure, computer analysis methods are used that take known structures into account as well.

The prediction of the secondary structures used here was performed using the *Phyre* system (**P**rotein **H**omology/analog**Y** **R**ecognition **E**ngine, http://www.sbg.bio.ic.ac.uk/phyre/ – see also Bennett-Lovsey et al., 2008). The models produced by *Phyre* are based on finding a sequence alignment to an already known structure, copying the coordinates and relabeling the residues according to the input sequence. Different prediction algorithms are used to generate the final prediction sequence.

All work for entering the sequences to *Phyre* and retrieving the prediction results was kindly done by Joachim Reichelt, Helmholtz Centre for Infection Research, department of Structural Biology.

The prediction of secondary structures in peptides is difficult at best and should be considered only as an *indication* of what might actually be a structural feature. Even with proteins, which usually present a more fixed structure than flexible peptides, predictions are usually only valid for 70–80 % of all cases. The prediction algorithms do not take (potential) disulfide bridges into account, although those could be added later manually in predicted 3D models in many cases. The predicted sequences were not precise enough, however, to make an addition of those disulfide bridges possible in most cases. This is mainly due to difficulties in structure prediction of short peptides, further analysis would require information on structure data e.g. by crystallization.

The information generated by Phyre in this analysis gives either

- **C** for a "random coil" formation of the respective amino acids,
- **E** for a straight "extended" structure of the amino acids, or
- **H** for a helical structure that the amino acids form.

The first comparison was for a number of ten clones randomly picked from the library, compared to ten completely random peptides. The goal was to determine if certain structural features were either dominant or preferred, or if any structural components would show up at all.

As can be seen in Figure 44, the variable sequence part of peptides from the library does not differ much from the fully randomized control sequences shown below. The

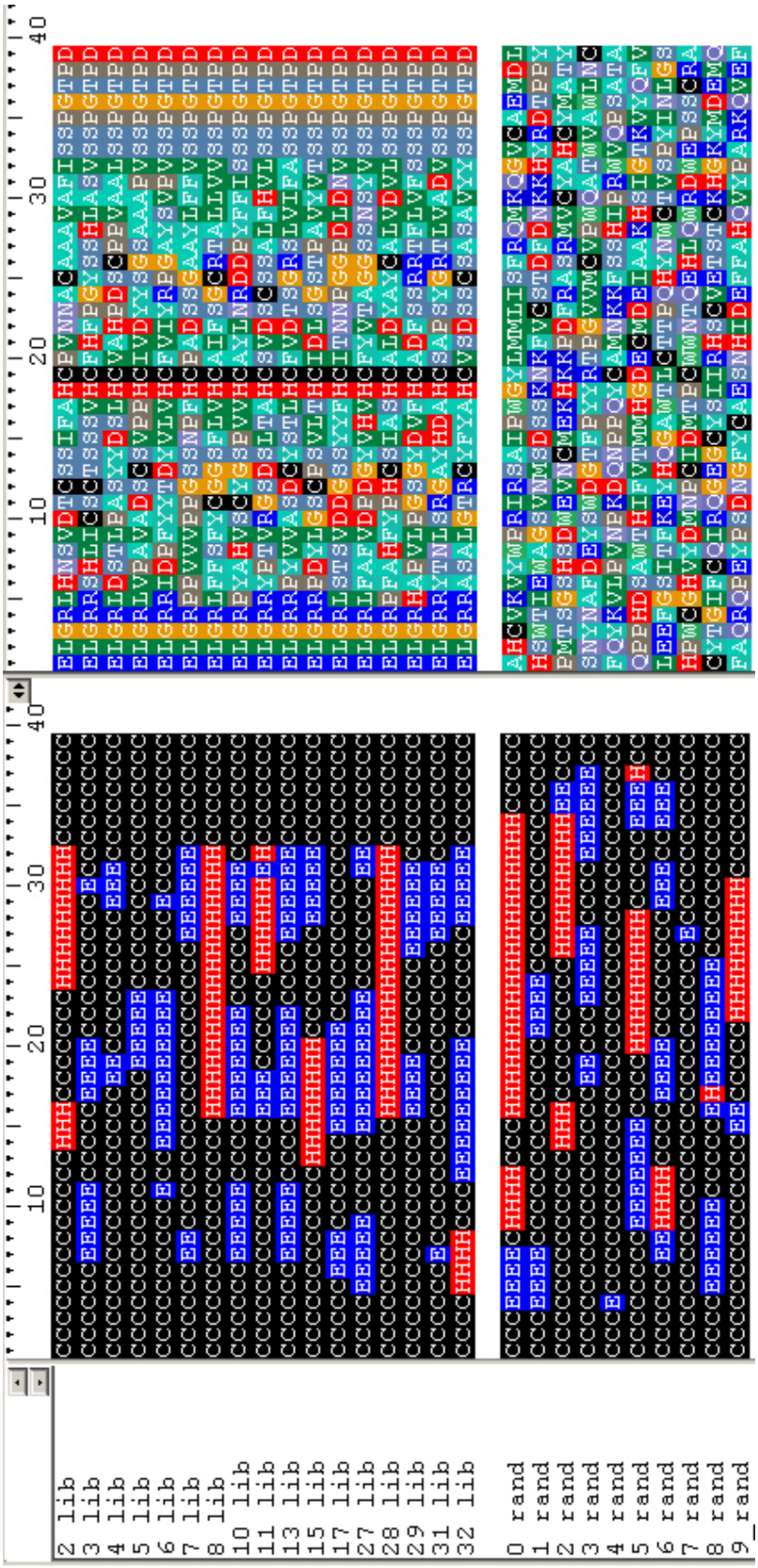


Figure 44: Secondary structure predictions for peptides
The first 17 peptides were randomly picked from the CPL19YS-2 library, while the last ten peptides have been randomly generated with each position allowing each of the 20 natural amino acids.
The first 4 amino acids of the library peptides are all the same (constant region of ELGR), as well as the positions 18 and 19 (HC) and the last seven amino acid positions (SSPTGPD, the linker to the pIII phage protein). The original amino acid sequences are shown right to the secondary structure predictions

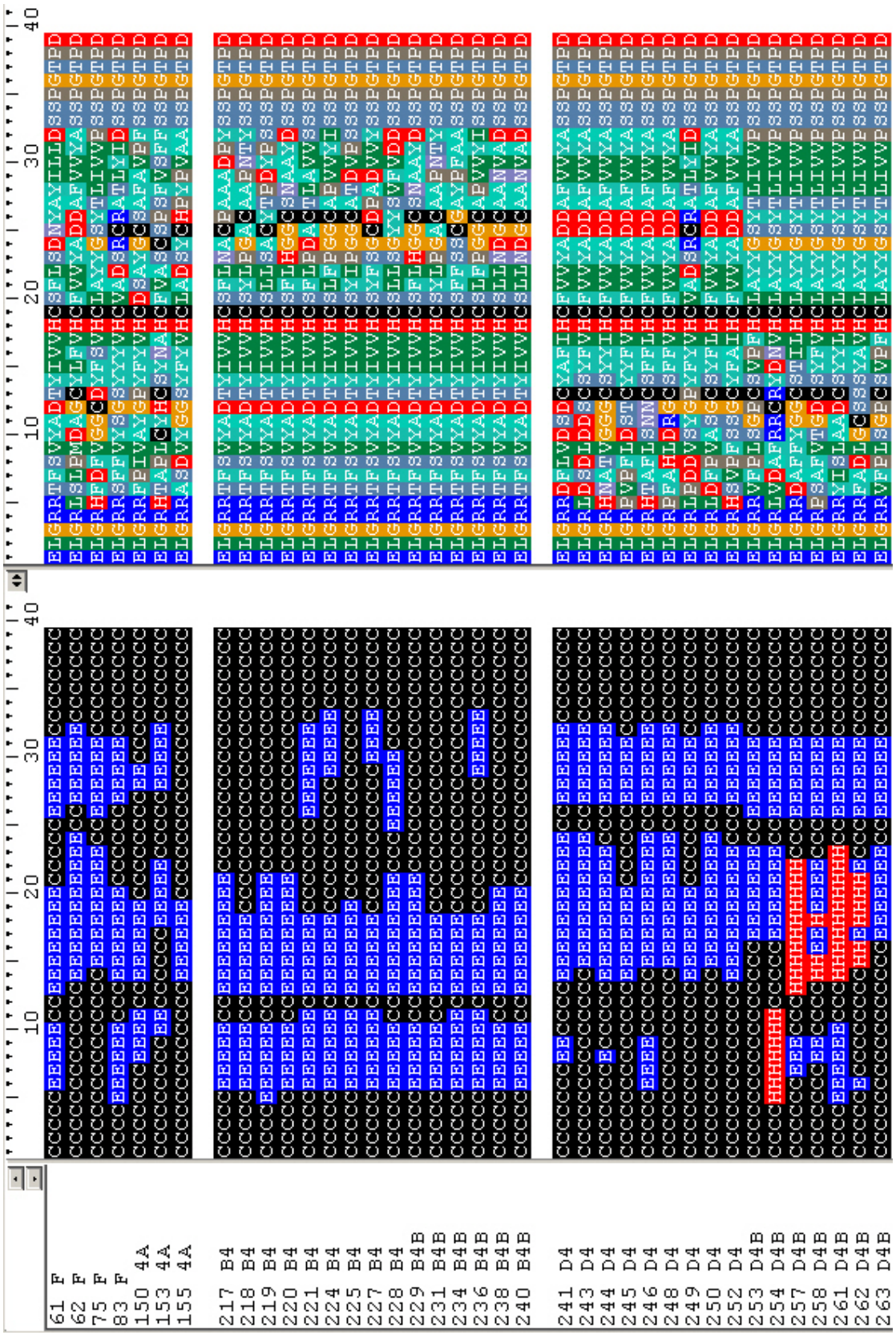


Figure 45: Secondary structure predictions for peptides selected on CD28
The first seven peptides were enriched from the CPL19YS-2 library in two pannings on CD28 (see Figure 33), the sequences 217–240 and 241–263 were enriched in the recombination experiments (section 3.3.3.4) from the CPL19YS-2-CD28B and CPL19YS-2-CD28D sublibraries, respectively. The first 4 amino acids of the peptides are all the same (constant region of ELGR), as well as the positions 18 and 19 (HC) and the last seven amino acid positions (SSPTGPD, the linker to the pIII phage protein). The original amino acid sequences are shown right to the secondary structure predictions

fixed sequence parts in front and after the variable parts seem not to contribute to other structural features, being just random coil elements that are shared amongst all candidates from the library.

Especially for the last seven amino acids this is desirable, since those amino acids were supposed to act as flexible linker and not be part of any potential interaction site.

Analyzing the peptides that have been enriched in the pannings on CD28 (Figure 45) offers additional information: Without differentiating between different panning experiments, it can be clearly seen that the occurrence of helical structures is significantly less frequent (13 %) than it was in the peptides picked randomly from the library (36 %) or the fully randomized peptides (50 %). Since helices form rather rigid structures, this is an indication that flexibility was required for the enriched peptides to attach to the CD28 target.

The first seven peptides in Figure 45 were enriched independently, but still share a few properties: A stretch in the middle of the peptide (5–11 aa) is always forming an extended structure, which is flanked by a few random coil elements and in most cases another extended structure to the left and to the right. The fixed positions to the left and the right of the variable sequence parts are again random coil, as concluded already from the random library peptides.

In the recombination experiment, the first four clones (61–83) had been recombined with the original CPL19YS-2 library and the resulting new libraries been panned on CD28 again, after which new clones were enriched (several of them shown in this figure, sequences 217–240 from recombination with the left peptide part kept fixed, sequences 241–263 from recombination with the right peptide part kept fixed). Even considering that the left or right part of the sequence was not just random, but taken from four distinct clones, the enriched peptides look surprisingly similar in their predicted secondary structures: The majority of recombined enriched peptides was consisting of two closely neighbored extended stretches of amino acids (a stretch in the center of the peptide of again 5–11 aa, and another extended section of 5–6 amino acids).

The only five peptides breaking that pattern are the candidates 254–262, where Phyre predicts potential helical structure elements. It can not be resolved finally if those are just artifacts from small homologies found in the prediction process, but it is notable that the only five peptides with helical structure element prediction have emerged in the panning of the recombined CPL19YS-2-CD28D library on the CD28 (after the change in immobilization method, indicated by the D4B, see section 3.3.3.1).

3.5 Analysis of target-peptide interactions

3.5.1 Surface plasmon resonance affinity measurements

With several clones enriched in three different panning experiments, more data was required on the actual absolute affinities of the encoded peptides to the CD28 target. This data should be obtained by Biacore measurements (section 2.2.5.5), immobilizing peptides on the surface of the sensor chip and using CD28 as analyte to determine the affinity constants of the respective peptide to CD28.

The peptides required for the interaction studies could not be obtained directly from the phagemid preparations (no pre-determined cleavage site available), and measurements of the whole phagemid particles presenting the peptides would have influenced the results potentially too much (due to possible interaction of phage surface and the peptides; small changes in the valency of presentation that could have a huge impact on the perceived avidity).

The peptides were synthesized chemically by Raimo Franke from the department of Chemical Biology at the Helmholtz Centre for Infection Research, Germany. The synthesis method was Fmoc/tert-butyl solid phase synthesis on a Multisyntech Syro peptide synthesizer. The synthesis products were purified in a preparative HPLC (high performance liquid chromatography) step and the fractions containing the expected molecule mass were determined by LC-MS (liquid chromatography-mass spectrometry).

The synthesized peptides (Figure 46) lack an additional GELG-sequence on their N-terminus. This region was constant in all peptides from the CPL19YS-2 library and not specific for single enriched clones, therefore it was left out due to encountered difficulties in the chemical synthesis. A C-terminal lysine was added after a constant linker region SSPG. This lysine is biotinylated (ordered from *Novabiochem*) to allow immobilization of the peptides on streptavidin, fixing the peptide similarly to the original fusion to the pIII phage protein.

Problems encountered in the synthesis process, due to peptide length and aa composition, led to broad peaks in the chromatograms of HPLC and MS-LC. Aggregation is a probable explanation, considering the hydrophobic sequence stretches in the synthesized peptides. It was most probably also causing most of the synthesis attempts not to succeed, in that analysis of synthesis products did often not show the expected product mass at all.

Due to the limitations mentioned above, only 4 of originally 16 peptides specified were finally synthesized.

To determine if the recombination process was yielding candidates with improved affinity, recombined clones (#253, #254 and #261) were compared to a clone that had been

enriched in the first panning (clone #75) and that had been used for the recombination. This conferred a comparison of 13 variable amino acids on the amino-terminus side of the peptide — all showing the same 13 amino acids plus a SSPG linker at the carboxy-terminus. All had the potential for forming a single disulfide bridge within the variable peptide.

With the sensor chip consisting of four flow cells, the first cell was left empty (with no ligand coupled to the surface) as reference cells.

Peptides #75, #254 and #253 were picked for the measurements first (since there were only three cells left for measurements and peptide #261 had the highest grade of impurity from the synthesis).

The peptides were immobilized on the sensor chip as described in section 2.2.5.5 (see Figure 48).

The injection of the CD28 analyte, however, did not yield any difference in relative response (RU) for the flow cells with the immobilized peptides (fc2–fc4) from the control cell (fc1). The analyte was added first at a 100 nM, then at a 1 µM concentration (Figure 47).

Dissociation was detectable right after the injection of CD28 was ended, and the abrupt drop was the same for all four measured cells, not significantly different from the reference cell (with the reference cell giving the strongest signal). The dissociation did not follow a characteristic dissociation curve as expected from strong interaction partners. The signal did not return to the baseline, though, even after attempts to regenerate the sensor chip surface with 1.5 M NaCl, 3 M NaCl and 10 mM glycine (pH 3) (see Figure 49). This indicates that some molecules must have remained on the surface of the chip, regardless of immobilized peptide in a flow cell or not.

Figure 46: Sequences of peptides synthesized for Biacore affinity measurements in regard to CD28
The first peptide (#75) was isolated in the first panning experiments of the CPL19YS-2 library on the CD28 target, while the other three peptides were enriched in a panning of the recombined CPL19YS-2-D library on the CD28 target. They share the right sequence part with the peptide #75 due to the recombination.
While the phage displayed peptides had an additional GELG-sequence on their N-terminus, this region constant in all clones was left out due to synthesis difficulties. The C-terminal lysine added after the constant linker region SSPG is biotinylated to allow immobilization of the peptides on streptavidin, fixing the peptide similarly to the original fusion to the pIII phage protein. In the library peptides, the pIII protein is fused to the SSPG linker after an additional TPDI sequence (see p.59, Figure 16).

Furthermore, after injection of the CD28 analyte solution into a flow cell, no additional peptide could be immobilized on the streptavidin chip (data not shown here), even using much larger amounts of peptides than in the initial immobilization step. The CD28 protein itself was ruled out as binding irreversibly to the surface, since the "background signal" remaining after injection of CD28 was below 100 RU, while a molecule of a mass of ~42 kDa (mass of reduced monomer of the CD28/fc fusion) completely covering the sensor chip surface would yield signals of several thousand RU.

While biotin, as a rather small molecule (244 Da), could explain the shift in mass as well as the inaccessibility of the surface to biotinylated peptides, the manufacturer of the CD28/fc fusion denied remaining traces of biotin in the final preparation of the protein, and the PBS buffer solution alone was not found to block the chip surface before.

The experiment was repeated with another sensor chip to verify the measurements, giving similar results.

Injection of CD28 into the flow cells of a Biacore CM5 chip (same surface as the streptavidin SA chip, just without the covalently bound streptavidin) resulted in a temporary binding to the surface (as expected), and returned to the base line after injection. This shows that the observed phenomenon of irreversible binding to the chip surface is dependent on the streptavidin and its interaction with some substance in the CD28/fc solution, although this substance could not be determined.

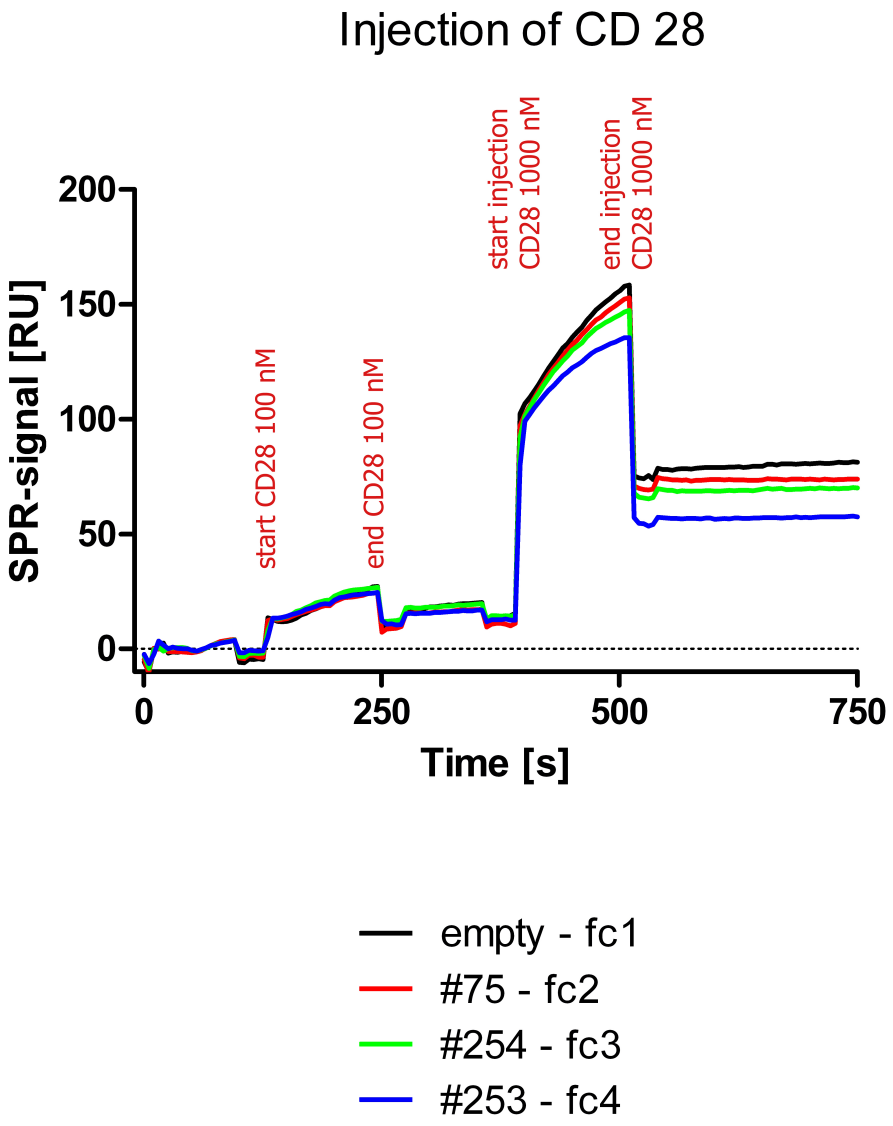


Figure 47: Injection of CD28 analyte to measure the affinities of peptides immobilized on a streptavidin SA *Biacore* chip
CD28 was injected into the four flow cells for 120 s, at a flow rate of 30 µ. Two different concentrations of CD28 were used.

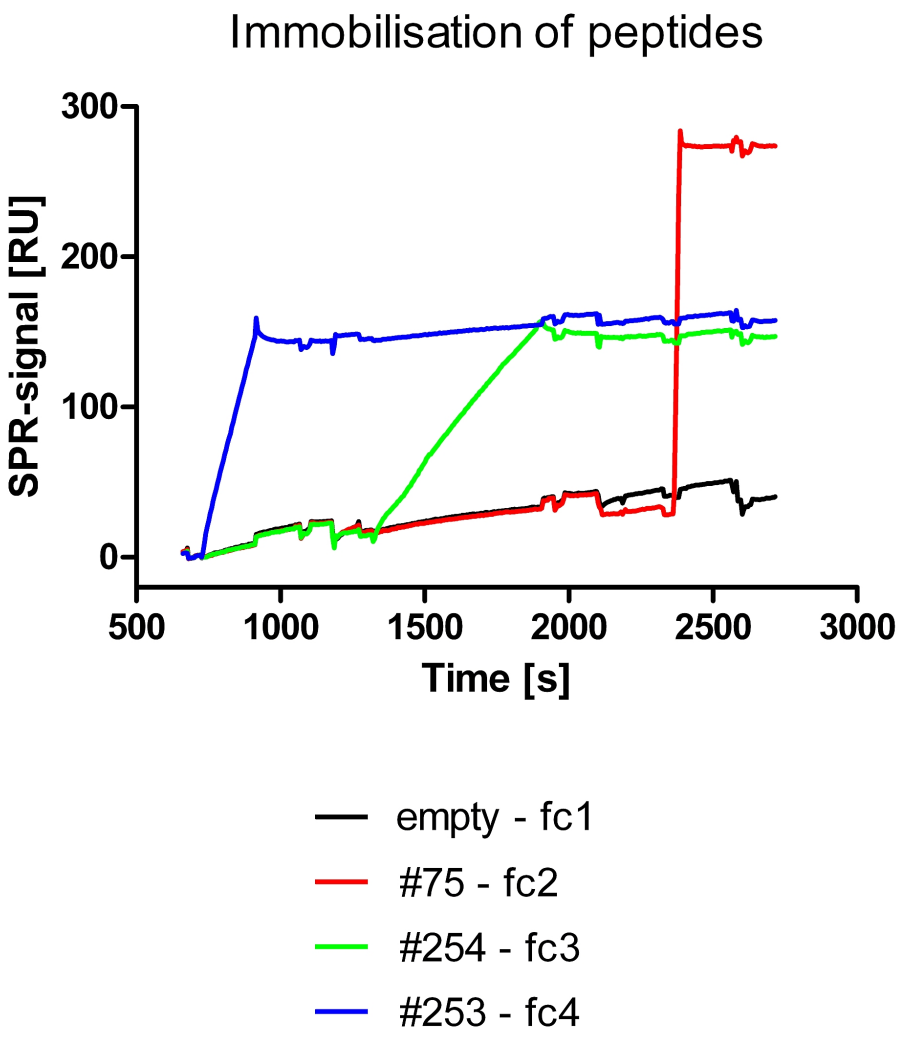


Figure 48: Immobilization of three selected peptides on a streptavidin SA *Biacore* chip
Three different peptides were immobilized on the Biacore chip in respective flow cells (fc2–fc4), the first flow cell (fc1) was left empty as a reference cell. Peptide was added until RU values between 150 and 300 were reached to allow for sensitive measurements later after adding of the analyte.

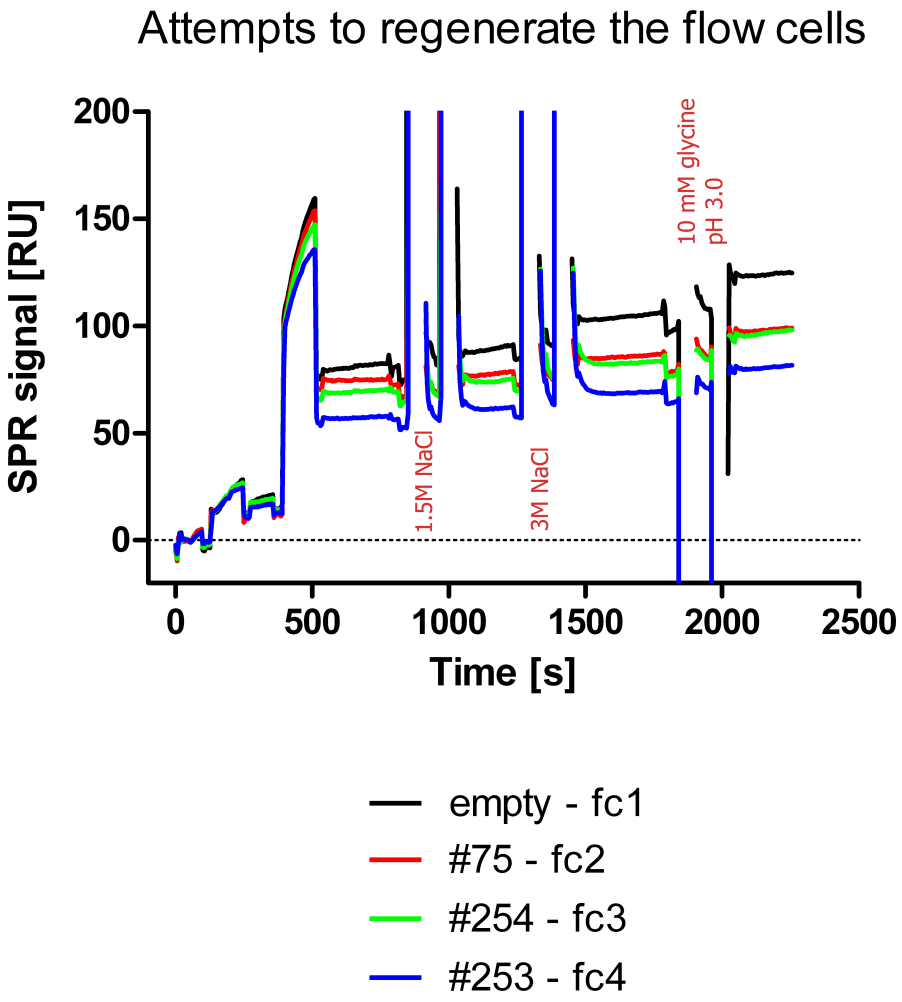


Figure 49: Attempts to regenerate the flow cells of a streptavidin SA *Biacore* chip after injection of CD28
The flow cells on the Biacore chip used in the experiment in Figure 47 were attempted to regenerate in three different steps after the affinity measurements that were conducted until second 500. Regeneration of the chip surface was attempted by rinsing the flow cells twice for one minute each at a flow rate of 30 µL/min, using 1.5 M NaCl, 3 M NaCl and 10 mM glycine (pH 3).

3.5.2 Relative ranking of peptide affinity by phage ELISA

A phage ELISA experiment, comparing the affinity of defined amounts of different phagemid particles presenting encoded peptides to the CD28 target, allows an estimation of the relative ranking of the peptide candidates.

It still must be taken into account that differences in fusion protein expression levels, potential differences in valency and the structural environment of the phage particle surface can influence the results.

For the relative ranking, the clones selected in the Biacore measurements (clones #75, #253, #254 and #261, section 3.5.1) were packaged into phagemid particles and the tetracycline/kanamycin titers (amount of helper phage) as well as the tetracycline/ampicillin titers (amount of phagemid particles) determined.

While a ratio of 1:10 to 1:5 was expected (and achieved with other packaged phagemid particles), these packaged clones had much higher ratios of M13K07 helper phage to phagemid particles of between 20:1 and 500:1. Their ratio of native pIII to fusion-protein pIII was not known, but could be different from the ~10 % presentation value determined in section 3.3.1 due to the observed difference in the ratio of helper phage to phagemid particles.

In the actual ELISA measurements, ~10^9 phage particles (from each clone) were incubated with the immobilized CD28 target in separate microtiter wells as described in section 2.2.5.6.

However, there was no observable difference at all between the amounts of phage remaining bound to the microtiter plate wells after washing, be it M13K07 or a fraction from the original CPL19YS-2 library (both as controls) or the four clones of interest (clones #75, #253, #254 and #261).

Changes in washing stringency or repeating the experiment gave similar results of no observable increased affinity of the phagemid particles encoding the enriched peptides.

3.5.3 Conclusions from affinity measurements

The enrichment experiments had shown results indicative of successful enrichment of high affinity ligands; panning experiments from the literature using primarily monovalently displayed peptides have never been found to enrich ligands with K_D values worse than ~1500 nM.

Although the measurements were hindered by the observed unexpected irreversible binding to the sensor chip surface after injection of the CD28 analyte solution, an affinity value of better than 500 nM would have been noticeable using analyte concentrations of up to 1 µM. So while the measurement itself showed to be not optimized methodically,

a high affinity interaction between the selected synthetic peptides and the CD28 target protein could not be quantified.

Since both experiments that were performed to assess the affinities of the peptides did not give any indication of high affinities, their implication of low affinity of the enriched peptides toward CD28 should not be overlooked. But these implications are contradicting the results of the panning enrichments, and both experiments could have been impaired by several factors:

The plasmon resonance measurement was performed using synthetic peptides. While it could not be guaranteed that these peptides would show the same "folding" (if it can be expected from the peptides used here), the oxidative conditions used to allow for the forming of intramolecular disulfide bridges could not be guaranteed to match the cell environment as well. The purity of the synthetic peptides was also not as high as one could have desired, although this was due to the difficult synthesis process which yielded only minor amounts of peptide. The effects of by-products or substances potentially remaining from the synthesis could also affect the peptides' stability or their potential to bind to the CD28 target molecule, and no information was available if the immobilization procedure was presenting the peptides in a similar way to the conditions on the phage surface.

By using the ELISA measurement, the use of *synthetic* peptides should be avoided by measuring the affinity of peptides presented on the phage surface, matching the conditions used in the panning (synthetic peptides could have been added competingly in a second step if there had been detectable affinity of phagemid particles) and ruling out effects of the synthesis process. The observed difficulties in packaging the specific clones of interest were different from the conditions observed in the original panning experiments, though, and could not be corrected by repeated packaging preparations. If expression of the fusion peptide was impeded in these clones enriched specifically, it would explain non-observable affinity in the ELISA measurement as well.

Experiments to re-clone just the peptide encoding sequence into an expression vector appeared to be a good way to determine if the packaging problems were an effect of the encoded peptides itself (potentially causing growth disadvantages to host cells, either being toxic or inhibiting the metabolism) or acquired by mutation of a remote part of the phagemid DNA. These experiments could not be conducted due to time limitations, however, leaving the problems noticed, but unexplained.

4 Conclusions

4.1 Rationales for designing a peptide library

The aim of this work was to establish and evaluate a phage displayed peptide library that is suitable for the isolation and characterization of very high affinity ligands for variable targets, thus not being tailored for specific purposes.

Starting with the design of the library, a phage displayed peptide library was chosen. While today peptide libraries are mostly used for epitope mapping (e.g. Böttger and Böttger, 2009; Uchiyama et al., 2005) and antibodies/antibody fragments are used for the isolation of high affinity ligands especially for pharmaceutical targets (Dimitrov and Marks, 2009), there have also been numerous successes to identify high affinity peptides, e.g. a peptide with a K_D of 68 nM to glycoprotein GPIα (Benard et al., 2008), the model selection of a mRNA-displayed peptide binding to a monoclonal antibody against human TP53 (K_D=6 nM) by Shiratori et al. (2009) or a p53 competing peptide for the oncoprotein MDM2 (K_D=3.4 nM) by Pazgier et al. (2009).

Although antibodies are used often as pharmaceutical candidates (Stockwin and Holmes, 2003), they later need to be trimmed down in size as far as possible to reduce immunogenicity (Irving et al., 2001), which leads more researchers towards peptides as pharmaceutic candidates (Ladner et al., 2004).

Peptides possess several more intrinsic advantages over antibodies as drug candidates, including lower manufacturing costs (synthetic versus recombinant production, see Ladner et al., 2004), a higher activity per mass (15–60-fold, assuming 75 kDa for one combining site of an antibody and 10–50 amino acids), less chance of unintended interaction with the immune system (assuming the peptide contains no known immune-system signaling sequence, Ladner et al. (2004)), and better organ or tumor penetration (Graff and Wittrup, 2003; Ruoslahti et al., 2005). Therefore, the small molecules obtained from peptide libraries offer new approaches and are favored by the pharmaceutic industry if available.

Considering these facts, the selection of a phage displayed peptide library was a viable choice for the library format established in the course of this thesis.

4.2 Covering complexity in library design

In the design of peptide libraries the potential complexity of a library is an important quality criterium. However, the higher the potential complexity is, the lower is the fraction of this complexity which it will be possible to cover in actually making the library.

Even with new display technologies as e.g. *in vitro*-compartmentalization (section 1.2.1.4), libraries are limited in size such that at the most only fully randomized 9- to 10-mers could be adequately represented (Matsuura and Yomo, 2006).

An additional problem arises then in the efficiency of selection of individual clones under panning conditions where maybe less than 10 displaying particles of an individual clone are present in the initial challenge. With individual concentrations that low, especially when long washing steps are included, the retention of clones in the first round becomes a "lottery" (Szardenings, 2003).

One response to this limitation could be deliberate reduction of the full complexity, whereby a reduced repertoire choice is self-imposed, or just tailoring towards specific targets, as with the ribonuclease inhibitor derived leucine-rich-repeat libraries of Stumpp et al. (2003) or β-barrel scaffold libraries like the GFP-based libraries developed by Abedi et al. (1998).

Reduction of complexity without reducing peptide length involves rational considerations on similarity of structure between different amino acids. Such model concepts should be confirmed experimentally. This was done e.g. done by Fellouse et al. (2004) who created a synthetic library using a Fab framework with CDR diversity restricted to four amino acid types (tyrosine, serine, alanine, aspartate), obtaining antibodies with 2–10 nM affinities to human vascular endothelial growth factor (VEGF). Further restriction to just tyrosine and serine (Fellouse et al., 2005) still yielded antibodies with 60 nM affinities to VEGF (see e.g. the review of Koide and Sidhu, 2009, for further details on reduction of amino acid diversity). This concept has rarely been used for peptide libraries (e.g. Larsson et al., 2002), but has never been tested thoroughly for peptide libraries (Szardenings, 2003).

The approach of a deliberately limited complexity has been taken in the library design described here, resulting in a reduced repertoire of amino acids at each position. These are arranged in a set of four cassettes which can be recombined with each other either during the process of creating the library or at any later stage. The potential complexity of each cassette is below 10^8. The library was assessed with regard to its constitution and expression of the fusion peptide, thus a library comprising $> 2 \times 10^8$ clones should carry all possible cassettes according to the library design (as in Figure 16, p.59).

4.3 Recombination as a key feature of the CPL19YS-2 library

The CPL19YS-2 library consists of four cassettes that would yield a theoretical complexity of 1.6×10^{25} when fully recombined. This represents a much higher potential complexity than what is common in conventional peptide libraries.

While conventional phage libraries are limited in their maximal complexity due to the above mentioned low concentrations per clone in the initial challenge round, the full *potential* complexity of the CPL19YS-2 library is not available in a single panning round as well. The amount of different clones that can be screened per panning round is still in the 10^8 range, but new diversity can be generated easily between panning experiments, adding new combinations from the pool of the potential 1.6×10^{25} candidates.

Recombination as a tool to add diversity and additional complexity has been postulated already by Moore et al. (1997). Methods used to add library diversity like e.g. successive rounds of point mutations (Yu and Smith, 1996), will increase the diversity *linearly*. Recombination between already existing or induced mutations as already performed by e.g. Stemmer (1994a,b), in contrast, will yield *exponential* potentiation of the existing diversity.

While different DNA shuffling methods have been established since 1994, their requirement of ~70 % sequence homology at the recombination sites makes them only suitable for larger proteins where low density mutations are reassorted, not mostly completely randomized smaller peptide libraries where such a sequence fixation would cover a large part of the supposedly variable sequence.

Shuffling cassettes in an ordered way by the method of *Cosmix Plexing* was shown to be a feasible way to increase the complexity of the primary library and to generate new variants, giving access to a highly complex, but still accessible library of peptides. As already established by Collins et al. (2001), this method did not require the long homology stretches used in other recombination technologies. The recombination process used was not a one-time action but could be repeated as often as desired to generate new diversity.

The recombination technique *Cosmix-Plexing* can be used for three different purposes:

- **Increase of library diversity:** The clones from the primary CPL19YS-2 library were recombined with themselves, generating ~2.4×10^8 combinations in a single enzyme restriction, religation and cloning step (Section 3.2.2.3). This procedure was not observed to confer any bias on sequences recombined (section 3.2.2.5).

- **Backcrossing of enriched clones with the CPL19YS-2 library:** *Cosmix-Plexing* can also be used to "backcross" enriched clones with a library to increase the variety in defined cassettes, while keeping the enriched sequences of the remaining

cassettes (Figure 12b). Using this approach, it is possible to screen the library for sequences that are synergistic to the sequences in the kept cassettes. It can also be analyzed which cassette(s) of an enriched peptide are contributing most to enrichment by replacing specifically either one or another sequence part.

This method has been used with clones obtained from a panning on CD28 (section 3.3.3.4). The newly generated sublibraries CPL19YS-2-CD28B and CPL19YS-2-CD28D showed a much faster enrichment and earlier increase of the output:input in the panning cycle than the original CPL19YS-2 library, indicating that the new sublibraries possessed already higher target-specific affinities at the beginning of the panning experiment. It was also notable that certain pre-selected cassettes were enriched over others, giving leads for further analysis of interacting motifs.

- **Recombination between enriched clones:** Clones from preselected populations (e.g. $>10^4$), as the clones obtained after 1–2 panning rounds, often contain several different motifs that confer target specific affinity. If these clones are recombined using *Cosmix-Plexing*, a new library with increased *target-specific* diversity ($>10^8$ clones) could be generated and be subjected to further panning rounds. This has not been applied during this thesis, so further testing is required to confirm the theoretical advantage of this technique.

The recombination process has been established and shown to be working without any noticeable bias introduced by the recombination process.

4.4 Considerations for secondary structure in the CPL19YS-2

Guided development of secondary structure in the CPL19YS-2

Working with peptide libraries to identify highly affine ligands, one encounters their limits when the target requires more than just linear chains of amino acids. Linear peptide motifs will be as they are most common in short peptide libraries (<8–12 aa, as are the lengths of peptides in commonly used libraries — see e.g. Noy et al., 2008) or in any clones that do not form stable secondary structure.

While it is possible to isolate short peptides that assume non-linear conformations at least when bound to specific targets, as for example a hexamer mimicking a nicotinic acetylcholine receptor site (by Balass et al., 1993), it is generally accepted that longer peptides increase the chance to form secondary structures: The longer peptides get, the more chance there is for intra-molecular stabilization in the form of hydrogen-bonds

or hydrophobic sidechain interactions. These restrictions reduce entropy, and they also restrict the available conformations a peptide may take. Longer peptides allow for further stabilization by the potential for formation of alpha helices or beta sheets. These motifs usually take at least ten to fifteen amino acids to form, which is already more than most peptide libraries comprise. Short polypeptides generally do not exhibit much of these secondary structure in solution, as the entropic cost associated with the folding of the polypeptide chain is not compensated for by a sufficient increase in stabilizing interactions (Hudgins and Jarrold, 1999).

These deliberations were taken into consideration in designing the CPL19YS-2 library, in particular by generating a library with a variable display segment of 33 aa size which has a strong preference to form disulfide bridges at different positions. The loops potentially formed by these bridges can vary in length (between 4 and 8 aa length). It has been reported by others that a constraining cysteine loop often requires optimization in its loop length (Nagi and Regan, 1997). This variable element in the library should allow selection of the "optimized disulfide loop length" already during the early screening steps.

Analyzing the clones from the CPL19YS-2 that were isolated either randomly or from enrichments during different pannings, the observations of Kay et al. (1993) with respect to counter-selection of unpaired cysteine residues selected against were not observed. The randomly isolated clones, without any previous selection, showed a random distribution of 0, 1, 2 or 3 cysteines in addition to the fixed central cysteine. Although this distribution was shifted towards more cysteines, due to an observed bias in the synthesis of the oligonucleotides used to construct the library, there was no indication of a bias against unpaired cysteines as observed by Kay et al. (1993). it should be noted that their phage-displayed library presented the variable peptide on five copies of the phage pIII protein, requiring correct folding and function (as pIII) of the hybrid protein. Our library is preferentially monovalent, i.e. only one hybrid pIII protein is present on average per two phage particles. In CPL19YS-2 there are still some four functional non-hybrid pIII molecules available for functional attachment and injection of DNA from the phagemid particles.

While the original library had a preference (~50 % of all clones) for peptides with two cysteines, it is notable that the majority of clones enriched in different panning experiments (56 out of 62 – 90.3 %) possesses an even number of cysteines. This is significantly higher than the library average and therefore strongly supports a positive effect on enrichment from the formation of intramolecular disulfide bridges. It can not be concluded at this point if this positive effect is due to enhanced rigidity of the peptide in general, aggregating hydrophobic residues or stabilizing hydrogen bonds, or an improved presentation of an amino acid motif in the disulfide "loop". In the two experiments carried out with antibodies as targets (sections 3.3.2 and 3.3.5) where a linear motif

was expected, the presumed disulfide bonds in enriched clones were always outside the detected motif. These results support our arguments which put emphasis on the potential for generating disulfide bridges. Such an approach had previously also been postulated by e.g. McLafferty et al. (1993).

4.5 Application of the CPL19YS-2 library for various target screenings

Peptide libraries used in the screening for affinity ligands are either tailored for specific purposes, by limiting their amino acid composition or by using defining structural/scaffold elements, or alternatively were in contrast conceived as "universal" random libraries without any structural bias (Clackson and Wells, 1994). Both approaches have yielded successful enrichment of specific ligands on some of the examined targets where it is noted that this required optimization of varying protocol parameters in the screening process. It is also discussed that some targets may not be suitable for this methodology, as there are proteins whose interaction sites would have a low maximal affinity to small molecules (Cheng et al., 2007), which has been difficult to predict.

In the current work, different targets have been used to screen the CPL19YS-2 peptide library for affinity candidates. While it was possible to confirm the display and screenability on model targets (anti-p53 Ab-6 antibody, and later on by lab members on antibodies from a polyclonal serum) by the isolation of candidates with expected epitopes, other screening experiments could not enrich single phage clones at all (pannings on pure streptavidin beads, pannings on a part of the M3 protein of group A streptococci).

That the unsuccessful panning on streptavidin beads enriched for no specific motif was expected, as the library was designed to exclude the common streptavidin-binding HPQ motif. The library does not contain the integrin binding RGD motif either to reduce the potential of unspecific enrichment if the library is used for panning on eukaryote cells.

The unsuccessful screening on the M3 protein might well be due to the target being unsuitable for a peptide library. The M3 protein is interacting with collagen at several sites and remains in a flexible, extended state (Dinkla et al., 2007, 2009). This may be disadvantageous for a screening of a peptide library where the candidates are not large enough to be interacting with several collagen sites. M3 would not present a target pocket which could be expected to strongly bind a peptide ligand.

The different pannings on CD28, either with the original CPL19YS-2 library or recombined libraries to make use of the implemented recombination features, showed promising results in the enrichment behavior which led to expectations of high affinities. As these

were not confirmed in the peptide-target interaction analysis, it remains unclear if the affinities were too low to be measured, leaving the enrichment results as potential artifacts of phage matrix mediated binding of the peptides to the target (as noted e.g. by Webster, 1996), or if the correct affinity values could not be determined due to difficulties in the chemical synthesis of the examined peptides.

The products obtained from chemical synthesis led to broad peaks in the chromatograms of HPLC and MS-LC, instead of clearly defined peaks indicating pure products. Aggregation due to partially hydrophobic sequences could explain the encountered problems as well as incomplete synthesis that has been observed during the synthesis process.

The dependence on chemical synthesis in order to be able to characterize isolated peptides outside of the phage context must be seen as a design flaw in a library that is supposed to encode peptides which could not be easily provided by chemical synthesis.

Although high affinity ligands against many different targets have been obtained using phage display libraries displaying antibodies (or -fragments), there have been many considerations of how to improve these methods. Following the natural approach of shuffling the heavy and light chains of antibodies, recombination has been used to increase library complexity (de Haard et al., 1999; Griffiths et al., 1994; Marks et al., 1991; Sblattero et al., 2001; Zhang et al., 2007). These different methods of ordered recombination have been shown to be successful with antibodies, while recombination in peptides has not been used to this extent, mostly due to required homologous sequence stretches. Using Cosmix plexing (Collins et al., 2001) as recombination method, by including cycles of recombination between selection the effective diversity can be considerably increased, and the theoretical maximum diversity can be approached step by step. This approach is equivalent to V gene shuffling, but carried out in parallel, rather than in series, as it is performed in other methods (compare e.g. Sblattero et al., 2001).

4.6 Concluding remarks

In comparison to mutation strategies which allow linear increases in complexity, an exponential potentiation could be achieved by recombination between existing or induced mutations present in a population (Moore et al., 1997). While it was not possible to confirm here that this yielded higher affinity ligands than panning from the primary CPL19YS-2 library (as no affinities were measurable for the isolates from the panning on CD28), the process itself has been used to recombine enriched clones with the original library at desired positions in the peptides. Sequencing of these clones confirmed the enriched sequence parts as well as recombined sequence parts. Subsequent panning experiments with these clones indicated strong preferences for several conserved parts

of the original enriched sequences after recombination, allowing an easy mapping of sequences that conferred to positive selection. This could also be used as a tool to analyze which part of a peptide contributes most to affinity selection easily.

Nevertheless, the general behavior of the library in the panning experiments, showing increase in input:output ratios and generally observed enrichments of similar secondary structure patterns after panning on CD28 show that the library is useable for panning experiments requiring potentially complex ligand structures without prior tailoring of the library.

Consequently, the general usability of the CPL19YS-2 library for screening has been confirmed and is also being used in further screening experiments in our laboratory.

5 Outlook

In this work, a novel peptide library CPL19YS-2 was created, its quality was validated and quantitatively assessed. The essential recombination features have been shown to be easily available as tools to increase initial diversity of the library as well as for recombination later in/after a panning experiment.

With these results, a toolbox is accessible which allows the panning for medium length (longer than in most peptide libraries) and complex peptide structures as ligands for a broad range of possible binding partners. The absence of a preferential predefined structure as well as having high complexity provide a ready-to-use technology for panning on a wide range of targets.

Until now, the range of targets that have been chosen for panning does not yet allow conclusions as to the usefulness of generating ligands with more stable secondary structures. Further experiments in our lab have shown that enrichment from the library on targets different from the ones examined in this thesis showed better results, so the full effect of recombination needs to be assayed on these targets, further characterizing the impact this method could have on the affinity of selected peptides as is expected. Similar recombination work conducted on antibodies has already shown advantages (e.g. Sblattero et al., 2001), and its successful transferability to peptide libraries still needs further validation.

A limitation of the CPL19YS-2 library seemed to be in the potential difficulties arising when enriched identified peptide motifs should be analyzed. In order to measure the affinities of isolated motifs to their targets exactly, the ligand needs to be available without any effects from neighboring sequences and the phage context. While short peptides of less than ten amino acids can be synthesized chemically routinely, the peptide sequences from the CPL19YS-2 library presented large difficulties in the synthesis process, prohibiting affinity measurements for most of them.

If the library was to be modified, the cloning of a DNA sequence encoding protease cleavage site between the pIII protein of M13 and the encoded library would allow for the production and purification of single clones just from phagemid infected *E. coli* cultures rather than having to rely on chemical synthesis. The inclusion of an *amber* stop codon (TAG) between the displayed-peptide gene and the pIII gene would be another possibility. It would allow the production of the fusion protein for phage display in an

amber-suppressor strain (e.g. XL1-Blue from *Stratagene*), while a transformation of the phagemid into a non-suppressor *E. coli* strain could be used for expression and secretion of the library peptide.

Better results from the first panning round might be achieved by using *poly*valent display for just one round of panning (taking the low concentration of every single clone in the first panning round into account), returning to *mono*valent presentation for the following rounds, capturing potentially more of the medium- to high-affinity clones in the initial panning round. This could be achieved by either switching of the packaging system (e.g. *Hyperphage* from Rondot et al., 2001, which produces phage particles presenting only fusion proteins), or by using cell based systems that eliminate the use of helper phage completely. Such systems use bacterial cells for phage packaging that carry different phage gene encoding plasmids, allowing control over the valency of presentation (Chasteen et al., 2006), as used by Secco et al. (2009).

Taken together, this work gives access to a versatile peptide library of possible high structural complexity and presents first steps in its further characterization.

6 Summary/Zusammenfassung

6.1 Summary

Phage display of peptides is a technique that is used to express and present peptides as fusion proteins on the surface of a phage particle. These presenting particles are in general used to select for peptides which can interact with a given target molecule. The method is carried out most frequently in an empirical iterative affinity enrichment procedure starting with libraries comprising large numbers of variants.

In this work, a phage display peptide library was created (”CPL19YS-2“), which should possess the potential for formation of secondary structure in the library due to a length of 33 amino acids for the encoded peptides. This should allow for higher affinities to many various types of target molecules than that which is achieved by using common libraries of shorter peptides.

The library was designed with respect to the Cosmix-plexing method, which allows recombination of hyper-variable DNA regions. It uses type IIs restriction enzymes that recognize a defined DNA sequence, but cut in a fixed distance from that sequence. The enzyme can therefore bind outside of the hyper-variable region which encodes the library diversity. This method can be applied to increase the diversity of the primary library, but its main strength is to allow the recombination of different coding DNA cassettes of peptides from preselected populations which already showed affinity to a particular target, thus optimizing the binding capacities of these variants to this target.

The library was created using synthetic DNA and, together with a modified expression vector, assessed for its constitution. Furthermore, the Cosmix-plexing recombination method was validated and established for this library.

Expression and presentation of the library peptides were validated, and the suitability of the CPL19YS-2 library for screening was confirmed in affinity selection experiments.

In a screening of the CPL19YS-2 library for ligands binding to CD28 (a T cell surface protein), specific peptide variants were enriched and recombined using Cosmix-plexing. Indications of high affinity to CD28 (and of higher affinities for the recombined variants) could not be confirmed in BIAcore analysis.

The inability to obtain clear evidence of affinity of the synthetic peptides was assumed to be due to difficulties in the chemical synthesis of such peptides which contained some very hydrophobic stretches. Some candidates have still not been successfully synthesized chemically.

6.2 Zusammenfassung

Phage display von Peptiden stellt eine Technik dar, mit der Peptide an einem Oberflächenprotein eines Phagenpartikels als Fusionsproteine exprimiert und damit präsentiert werden können. Diese präsentierenden Phagenpartikel werden im Allgemeinen dazu verwendet, Peptide zu selektieren, die mit vorgegebenen Zielmolekülen interagieren. Diese Methoden wird in den meisten Fällen in Affinitäts-Anreicherungen mit iterativen Zyklen angewendet, wobei zu Beginn der Anreicherungen meist Bibliotheken mit einer hohen Anzahl verschiedener Peptide eingesetzt werden.

In dieser Arbeit wurde eine Phage Display-Peptid-Bibliothek erstellt ("CPL19YS-2"), die durch eine Peptid-Länge von 33 Aminosäuren das Potential zur Ausbildung von Sekundärstrukturen besitzen soll. Diese Eigenschaft soll höhere Affinitäten zu einer Vielzahl verschiedener Zielmoleküle zulassen, als es bisher mit kürzeren Peptiden aus gängigen Peptid-Bibliotheken möglich ist.

Die Bibliothek wurde mit Hinsicht auf die Methode des Cosmix-plexing entwickelt, da diese die Rekombination in hypervariablen DNA-Regionen erlaubt. Sie verwendet Restriktionsenzyme vom Typ IIs, die eine definierte DNA-Sequenz erkennen, aber in einem festgelegten Abstand von dieser schneiden. Dadurch kann die Bindung eines Restriktionsenzyms außerhalb der hypervariablen Region stattfinden, innerhalb welcher die Diversität der Bibliotheksstruktur kodiert ist. Mit Hilfe dieser Technik kann man die Diversität der Ausgangsbibliothek steigern, ihr größerer Vorteil liegt aber darin, dass sie die Rekombination verschiedener kodierender DNA-"Kassetten" von Peptiden aus vorselektierten Population erlaubt, die bereits eine Affinität für eine bestimmte Zieldomäne besitzen. Dadurch kann die Affinität der Varianten für ihr Zielmolekül optimiert werden.

Die mittels synthetischer DNA erstelle Bibliothek wurde zusammen mit einem modifizierten Expressionsvektor bzgl. ihrer Zusammensetzung untersucht und die Rekombinationsmethode des Cosmix-plexing validiert und etabliert.

Die Expression und Präsentation der Peptide wurden validiert, und die Eignung der CPL19YS-2 Bibliothek für Screenings wurde in Affinitätsanreicherungs-Experimenten bestätigt.

In einem Screening der CPL19YS-2-Bibliothek auf Liganden für CD28, ein Oberflächenprotein von T-Zellen, wurden spezifische Peptide angereichert und mittels Cosmix-plexing rekombiniert.

Die Hinweise darauf, dass diese Peptide hohe Affinitäten besitzen (und die rekombinierten Varianten höhere Affinitäten), konnten in BIAcore-Analysen nicht bestätigt werden.

Die Schwierigkeiten beim Nachweis der Affinität der synthetischen Peptide liegen vermutlich in Limitierungen bei der chemischen Synthese begründet, da diese Peptide einige sehr hydrophobe Abschnitte aufweisen. Einige der Peptide konnten bisher noch nicht chemisch synthetisiert werden.

Acknowledgements

My first thanks go to Prof. Dr. John Collins for the chance to work and learn in his laboratory at the Helmholtz Center for Infection Research, and the opportunity to do my PhD under his guidance, from which I profited in many ways.

I also want to thank Dr. Ronald Frank and Prof. Dr. Singh Chhatwal for taking part in my thesis committee, looking at my work from a more "external" angle of view.

There were also many enlightening discussions with the members of the Chemical Biology department, on official occasions as in the lunch time discussions.

I also want to express my thanks to Simone Nieswandt and Jonas Kügler with whom I shared some time in the lab and had some fruitful discussions. May your pannings be successful.

Sonja Schweinsberg (University of Kassel, department of Biochemistry - Germany) helped me not only doing Biacore measurements, but also offered me her hospitality and showed me quite a bit of Kassel.

Raimo Franke made heroic attempts to synthesize my wonderfully hydrophobic peptides for me. I wish there had been better measurement results to reward you.

Tschong-Hun Im from the department of Genome Analysis helped me to save *dozens* of hours by doing larger scale plasmid preparation work for me.

I am also very happy to have such a great group of friends which helped me during different stages of my thesis, whether those were scientific or not.

I want to give a particularly big thanks to Allex for dealing/coping with my moods that may have been quite "difficult" at times.

Last, but not least, a very special thanks goes to my good friend caffeine.

Thanks, Erik

7 Bibliography

Abedi, M. R., Caponigro, G., and Kamb, A. (1998). Green fluorescent protein as a scaffold for intracellular presentation of peptides. *Nucleic Acids Res.*, 26:623–630.

Adey, N. B. and Kay, B. K. (1996). Identification of calmodulin-binding peptide consensus sequences from a phage-displayed random peptide library. *Gene*, 169:133–134.

Adey, N. B., Mataragnon, A. H., Rider, J. E., Carter, J. M., and Kay, B. K. (1995). Characterization of phage that bind plastic from phage-displayed random peptide libraries. *Gene*, 156:27–31.

Arad, G., Hillman, D., Levy, R., and Kaempfer, R. (2001). Superantigen antagonist blocks Th1 cytokine gene induction and lethal shock. *J. Leukoc. Biol.*, 69:921–927.

Arad, G., Levy, R., Nasie, I., Hillman, D., Barash, U., Minis, A., Rotfogel, Z., and Kaempfer, R. (2009). The dimer interface of CD28 binds superantigens and mediates signaling for Th1 cytokine gene induction. *Submitted. Mol. Cell.*

Aruffo, A. and Seed, B. (1987). Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. *Proc. Natl. Acad. Sci. U.S.A.*, 84:8573–8577.

Balass, M., Heldman, Y., Cabilly, S., Givol, D., Katchalski-Katzir, E., and Fuchs, S. (1993). Identification of a hexapeptide that mimics a conformation-dependent binding site of acetylcholine receptor by use of a phage-epitope library. *Proc. Natl. Acad. Sci. U.S.A.*, 90:10638–10642.

Beck, E. and Zink, B. (1981). Nucleotide sequence and genome organisation of filamentous bacteriophages fl and fd. *Gene*, 16(1-3):35–58.

Benard, S. A., Smith, T. M., Cunningham, K., Jacob, J., DeSilva, T., Lin, L., Shaw, G. D., Kriz, R., and Kelleher, K. S. (2008). Identification of peptide antagonists to glycoprotein Ibalpha that selectively inhibit von Willebrand factor dependent platelet aggregation. *Biochemistry*, 47:4674–4682.

Bennett-Lovsey, R., Herbert, A., Sternberg, M., and Kelley, L. (2008). Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins*, 70:611–625.

Berg, J. M. (1988). Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 85:99–102.

Bertschinger, J., Grabulovski, D., and Neri, D. (2007). Selection of single domain binding proteins by covalent DNA display. *Protein Eng. Des. Sel.*, 20:57–68.

Bertschinger, J. and Neri, D. (2004). Covalent DNA display as a novel tool for directed evolution of proteins in vitro. *Protein Eng. Des. Sel.*, 17:699–707.

Bianchi, E., Folgori, A., Wallace, A., Nicotra, M., Acali, S., Phalipon, A., Barbato, G., Bazzo, R., Cortese, R., and Felici, F. (1995). A conformationally homogeneous combinatorial peptide library. *J. Mol. Biol.*, 247:154–160.

Binz, H. and Plückthun, A. (2005). Engineered proteins as specific binding reagents. *Curr. Opin. Biotechnol.*, 16:459–469.

Böttger, V. and Böttger, A. (2009). Epitope mapping using phage display Peptide libraries. *Methods Mol. Biol.*, 524:181–201.

Chasteen, L., Ayriss, J., Pavlik, P., and Bradbury, A. (2006). Eliminating helper phage from phage display. *Nucleic Acids Res.*, 34:e145.

Chen, L., MacMillan, A. M., Chang, W., Ezaz-Nikpay, K., Lane, W. S., and Verdine, G. L. (1991). Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry*, 30:11018–11025.

Cheng, A. C., Coleman, R. G., Smyth, K. T., Cao, Q., Soulard, P., Caffrey, D. R., Salzberg, A. C., and Huang, E. S. (2007). Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.*, 25:71–75.

Christmann, A., Walter, K., Wentzel, A., Krätzner, R., and Kolmar, H. (1999). The cystine knot of a squash-type protease inhibitor as a structural scaffold for Escherichia coli cell surface display of conformationally constrained peptides. *Protein Eng.*, 12:797–806.

Cirino, P., Mayer, K., and Umeno, D. (2003). Generating mutant libraries using error-prone PCR. *Methods Mol. Biol.*, 231:3–9.

Clackson, T. and Wells, J. (1994). In vitro selection from protein and peptide libraries. *Trends Biotechnol.*, 12:173–184.

Click, E. M. and Webster, R. E. (1997). Filamentous phage infection: required interactions with the TolA protein. *J Bacteriol*, 179(20):6464–71.

Collins, J. (1997). *Phage display*, volume 1 of *Ann. Reports in Combinatorial Chem. & Mol. Diversity*, pages 210–250. ESCOM science Publishers.

Collins, J., Horn, N., Wadenbäck, J., and Szardenings, M. (2001). Cosmix-plexing: a novel recombinatorial approach for evolutionary selection from combinatorial libraries. *Rev. Mol. Biotechnol.*, 74:317–338.

Combelles, C., Gracy, J., Heitz, A., Craik, D. J., and Chiche, L. (2008). Structure and folding of disulfide-rich miniproteins: insights from molecular dynamics simulations and MM-PBSA free energy calculations. *Proteins*, 73:87–103.

Cunningham, B. C. and Wells, J. A. (1997). Minimized proteins. *Curr. Opin. Struct. Biol.*, 7:457–462.

Cwirla, S., Peters, E., Barrett, R., and Dower, W. (1990). Peptides on phage: a vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. U.S.A.*, 87:6378–6382.

Cwirla, S. E., Balasubramanian, P., Duffin, D. J., Wagstrom, C. R., Gates, C. M., Singer, S. C., Davis, A. M., Tansik, R. L., Mattheakis, L. C., Boytos, C. M., Schatz, P. J., Baccanari, D. P., Wrighton, N. C., Barrett, R. W., and Dower, W. J. (1997). Peptide agonist of the thrombopoietin receptor as potent as the natural cytokine. *Science*, 276:1696–1699.

de Haard, H. J., van Neer, N., Reurs, A., Hufton, S. E., Roovers, R. C., Henderikx, P., de Bruïne, A. P., Arends, J. W., and Hoogenboom, H. R. (1999). A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies. *J. Biol. Chem.*, 274:18218–18230.

Dedman, J. R., Kaetzel, M. A., Chan, H. C., Nelson, D. J., and Jamieson, G. A. (1993). Selection of targeted biological modifiers from a bacteriophage library of random peptides. The identification of novel calmodulin regulatory peptides. *J. Biol. Chem.*, 268:23025–23030.

Dennis, M. S., Zhang, M., Meng, Y. G., Kadkhodayan, M., Kirchhofer, D., Combs, D., and Damico, L. A. (2002). Albumin binding as a general strategy for improving the pharmacokinetics of proteins. *J. Biol. Chem.*, 277:35035–35043.

Devlin, J., Panganiban, L., and Devlin, P. (1990). Random peptide libraries: a source of specific protein binding molecules. *Science*, 249:404–406.

Dimitrov, D. S. and Marks, J. D. (2009). Therapeutic antibodies: current state and future trends — is a paradigm change coming soon? *Methods Mol. Biol.*, 525:1–27.

Dinkla, K., Nitsche-Schmitz, D. P., Barroso, V., Reissmann, S., Johansson, H. M., Frick, I. M., Rohde, M., and Chhatwal, G. S. (2007). Identification of a streptococcal octapeptide motif involved in acute rheumatic fever. *J. Biol. Chem.*, 282:18686–18693.

Dinkla, K., Talay, S. R., Mörgelin, M., Graham, R. M., Rohde, M., Nitsche-Schmitz, D. P., and Chhatwal, G. S. (2009). Crucial role of the CB3-region of collagen IV in PARF-induced acute rheumatic fever. *PLoS ONE*, 4:e4666.

Dower, W., Miller, J., and Ragsdale, C. (1988). High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res.*, 16:6127–6145.

Dybwad, A., Bogen, B., Natvig, J., Førre, O., and Sioud, M. (1995). Peptide phage libraries can be an efficient tool for identifying antibody ligands for polyclonal antisera. *Clin. Exp. Immunol.*, 102:438–442.

Eckert, D. M., Malashkevich, V. N., Hong, L. H., Carr, P. A., and Kim, P. S. (1999). Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell*, 99:103–115.

Endemann, H. and Model, P. (1995). Location of filamentous phage minor coat proteins in phage and in infected cells. *J Mol Biol*, 250(4):496–506.

Fellouse, F. A., Li, B., Compaan, D. M., Peden, A. A., Hymowitz, S. G., and Sidhu, S. S. (2005). Molecular recognition by a binary code. *J. Mol. Biol.*, 348:1153–1162.

Fellouse, F. A., Wiesmann, C., and Sidhu, S. S. (2004). Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc. Natl. Acad. Sci. U.S.A.*, 101:12467–12472.

Fukuda, I., Kojoh, K., Tabata, N., Doi, N., Takashima, H., Miyamoto-Sato, E., and Yanagawa, H. (2006). In vitro evolution of single-chain antibodies using mRNA display. *Nucleic Acids Res.*, 34:e127.

Fung, A., Li, P., Godoy-Ruiz, R., Sanchez-Ruiz, J. M., and Muñoz, V. (2008). Expanding the realm of ultrafast protein folding: gpW, a midsize natural single-domain with alpha+beta topology that folds downhill. *J. Am. Chem. Soc.*, 130:7489–7495.

Gallop, M., Barrett, R., Dower, W., Fodor, S., and Gordon, E. (1994). Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *J. Med. Chem.*, 37:1233–1251.

Gersuk, G. M., Corey, M. J., Corey, E., Stray, J. E., Kawasaki, G. H., and Vessella, R. L. (1997). High-affinity peptide ligands to prostate-specific antigen identified by polysome selection. *Biochem. Biophys. Res. Commun.*, 232:578–582.

Graff, C. P. and Wittrup, K. D. (2003). Theoretical analysis of antibody targeting of tumor spheroids: importance of dosage for penetration, and affinity for retention. *Cancer Res.*, 63:1288–1296.

Griffiths, A. and Tawfik, D. (2003). Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J.*, 22:24–35.

Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., Kontermann, R. E., Jones, P. T., Low, N. M., and Allison, T. J. (1994). Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.*, 13:3245–3260.

Groth, A. and Calos, M. (2004). Phage integrases: biology and applications. *J. Mol. Biol.*, 335:667–678.

Hackett, S. P. and Stevens, D. L. (1993). Superantigens associated with staphylococcal and streptococcal toxic shock syndrome are potent inducers of tumor necrosis factor-beta synthesis. *J. Infect. Dis.*, 168:232–235.

Hanes, J. and Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U.S.A.*, 94:4937–4942.

Harrison, J., Williams, S., Winter, G., and Nissim, A. (1996). Screening of phage antibody libraries. *Meth. Enzymol.*, 267:83–109.

He, M. and Taussig, M. J. (2002). Ribosome display: cell-free protein display technology. *Brief Funct Genomic Proteomic*, 1:204–212.

Holliger, P. and Riechmann, L. (1997). A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. *Structure*, 5:265–275.

Houston, M., Wallace, A., Bianchi, E., Pessi, A., and Hodges, R. (1996). Use of a conformationally restricted secondary structural element to display peptide libraries: a two-stranded alpha-helical coiled-coil stabilized by lactam bridges. *J. Mol. Biol.*, 262:270–282.

Hudgins, R. and Jarrold, M. (1999). Helix Formation in Unsolvated Alanine-Based Peptides: Helical Monomers and Helical Dimers. *J. Am. Chem. Soc.*, 121:3494–3501.

Iannolo, G., Minenkova, O., Gonfloni, S., Castagnoli, L., and Cesareni, G. (1997). Construction, exploitation and evolution of a new peptide library displayed at high density by fusion to the major coat protein of filamentous phage. *Biol. Chem.*, 378:517–521.

Irving, R. A., Coia, G., Roberts, A., Nuttall, S. D., and Hudson, P. J. (2001). Ribosome display and affinity maturation: from antibodies to single V-domains and steps towards cancer therapeutics. *J. Immunol. Methods*, 248:31–45.

Jönsson, U., Fägerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., Löfås, S., Persson, B., Roos, H., and Rönnberg, I. (1991). Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. *BioTechniques*, 11:620–627.

Karlsson, G., Jensen, A., Stevenson, L., Woods, Y., Lane, D., and Sørensen, M. (2004). Activation of p53 by scaffold-stabilised expression of Mdm2-binding peptides: visualisation of reporter gene induction at the single-cell level. *Br. J. Cancer*, 91:1488–1494.

Kay, B., Adey, N., He, Y., Manfredi, J., Mataragnon, A., and Fowlkes, D. (1993). An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene*, 128:59–65.

Kimura, R. H., Levin, A. M., Cochran, F. V., and Cochran, J. R. (2009). Engineered cystine knot peptides that bind alphavbeta3, alphavbeta5, and alpha5beta1 integrins with low-nanomolar affinity. *Proteins*.

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wölle, J., Plückthun, A., and Virnekäs, B. (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.*, 296:57–86.

Koide, S. and Sidhu, S. S. (2009). The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins. *ACS Chem. Biol.*, 4:325–334.

Ladner, R. C. (1995). Constrained peptides as binding entities. *Trends Biotechnol.*, 13:426–430.

Ladner, R. C., Sato, A. K., Gorzelany, J., and de Souza, M. (2004). Phage display-derived peptides as therapeutic alternatives to antibodies. *Drug Discov. Today*, 9:525–529.

Laemmli, U. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227:680–685.

Larsson, O., Thormeyer, D., Asinger, A., Wihlén, B., Wahlestedt, C., and Liang, Z. (2002). Quantitative codon optimisation of DNA libraries encoding sub-random peptides: design and characterisation of a novel library encoding transmembrane domain peptides. *Nucleic Acids Res.*, 30:e133.

Lee, S., Choi, J., and Xu, Z. (2003). Microbial cell-surface display. *Trends Biotechnol.*, 21:45–52.

Linsley, P. and Ledbetter, J. (1993). The role of the CD28 receptor during T cell responses to antigen. *Annu. Rev. Immunol.*, 11:191–212.

Low, N., Holliger, P., and Winter, G. (1996). Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. *J. Mol. Biol.*, 260:359–368.

Lowman, H., Bass, S., Simpson, N., and Wells, J. (1991). Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry*, 30:10832–10838.

Lowman, H. B. (1997). Bacteriophage display and discovery of peptide leads for drug development. *Annu Rev Biophys Biomol Struct*, 26(NIL):401–24.

Malmqvist, M. (1993). Biospecific interaction analysis using biosensor technology. *Nature*, 361:186–187.

Marciano, D. K., Russel, M., and Simon, S. M. (2001). Assembling filamentous phage occlude pIV channels. *Proc Natl Acad Sci U S A*, 98(16):9359–64.

Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D., and Winter, G. (1991). By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.*, 222:581–597.

Matsuura, T. and Yomo, T. (2006). In vitro evolution of proteins. *J. Biosci. Bioeng.*, 101:449–456.

Mattheakis, L. C., Bhatt, R. R., and Dower, W. J. (1994). An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. U.S.A.*, 91:9022–9026.

McConnell, S., Uveges, A., Fowlkes, D., and Spinella, D. (1996). Construction and screening of M13 phage libraries displaying long random peptides. *Mol. Divers.*, 1:165–176.

McLafferty, M., Kent, R., Ladner, R., and Markland, W. (1993). M13 bacteriophage displaying disulfide-constrained microproteins. *Gene*, 128:29–36.

Menendez, A. and Scott, J. K. (2005). The nature of target-unrelated peptides recovered in the screening of phage-displayed random peptide libraries with antibodies. *Anal. Biochem.*, 336:145–157.

Model, P. and Russel, M. (1988). *Filamentous bacteriophage*, volume 2 of *The bacteriophages*, pages 375–456. Plenum Publishers.

Moore, J. C., Jin, H. M., Kuchner, O., and Arnold, F. H. (1997). Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. *J. Mol. Biol.*, 272:336–347.

Nagi, A. D. and Regan, L. (1997). An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des*, 2:67–75.

Neidigh, J. W., Fesinmeyer, R. M., and Andersen, N. H. (2002). Designing a 20-residue protein. *Nat. Struct. Biol.*, 9:425–430.

Neidigh, J. W., Fesinmeyer, R. M., Prickett, K. S., and Andersen, N. H. (2001). Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. *Biochemistry*, 40:13188–13200.

Nemoto, N., Miyamoto-Sato, E., Husimi, Y., and Yanagawa, H. (1997). In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett.*, 414:405–408.

Neylon, C. (2004). Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res.*, 32:1448–1459.

Nguyen, A. and Daugherty, P. (2003). Production of randomly mutated plasmid libraries using mutator strains. *Methods Mol. Biol.*, 231:39–44.

Nord, K., Gunneriusson, E., Ringdahl, J., Ståhl, S., Uhlén, M., and Nygren, P. (1997). Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat. Biotechnol.*, 15:772–777.

Novotny, C. and Lavin, K. (1971). Some effects of temperature on the growth of F pili. *J. Bacteriol.*, 107:671–682.

Noy, K., Kalisman, N., and Keasar, C. (2008). Prediction of structural stability of short beta-hairpin peptides by molecular dynamics and knowledge-based potentials. *BMC Struct. Biol.*, 8:27.

Oldenburg, K., Loganathan, D., Goldstein, I., Schultz, P., and Gallop, M. (1992). Peptide ligands for a sugar-binding protein isolated from a random peptide library. *Proc. Natl. Acad. Sci. U.S.A.*, 89:5393–5397.

Olsen, M., Iverson, B., and Georgiou, G. (2000). High-throughput screening of enzyme libraries. *Curr. Opin. Biotechnol.*, 11:331–337.

Panayotou, G., Waterfield, M., and End, P. (1993). Riding the evanescent wave. *Curr. Biol.*, 3:913–915.

Pazgier, M., Liu, M., Zou, G., Yuan, W., Li, C., Li, C., Li, J., Monbo, J., Zella, D., Tarasov, S. G., and Lu, W. (2009). Structural basis for high-affinity peptide inhibition of p53 interactions with MDM2 and MDMX. *Proc. Natl. Acad. Sci. U.S.A.*, 106:4665–4670.

Pisabarro, M. and Serrano, L. (1996). Rational design of specific high-affinity peptide ligands for the Abl-SH3 domain. *Biochemistry*, 35:10634–10640.

Plückthun, A. and Pack, P. (1997). New protein engineering approaches to multivalent and bispecific antibody fragments. *Immunotechnology*, 3:83–105.

Powell, B., Rivas, M., Court, D., Nakamura, Y., and Turnbough, C. (1994). Rapid confirmation of single copy lambda prophage integration by PCR. *Nucleic Acids Res.*, 22:5765–5766.

Rakonjac, J., n Feng, J., and Model, P. (1999). Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. *J Mol Biol*, 289(5):1253–65.

Riechmann, L. and Winter, G. (2000). Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc. Natl. Acad. Sci. U.S.A.*, 97:10068–10073.

Roberts, B., Markland, W., Ley, A., Kent, R., White, D., Guterman, S., and Ladner, R. (1992). Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc. Natl. Acad. Sci. U.S.A.*, 89:2429–2433.

Roberts, R. and Szostak, J. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 94:12297–12302.

Rondot, S., Koch, J., Breitling, F., and Dübel, S. (2001). A helper phage to improve single-chain antibody presentation in phage display. *Nat. Biotechnol.*, 19:75–78.

Röttgen, P. and Collins, J. (1995). A human pancreatic secretory trypsin inhibitor presenting a hypervariable highly constrained epitope via monovalent phagemid display. *Gene*, 164:243–250.

Ruoslahti, E. (1996). RGD and other recognition sequences for integrins. *Annu. Rev. Cell Dev. Biol.*, 12:697–715.

Ruoslahti, E., Duza, T., and Zhang, L. (2005). Vascular homing peptides with cell-penetrating properties. *Curr. Pharm. Des.*, 11:3655–3660.

Russel, M., Clackson, T., and Lowman, H. (2004). *Introduction to phage biology and phage display*, pages 1–26. Phage display. Oxford University Press.

Russell, R. B. (1998). Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. *J. Mol. Biol.*, 279:1211–1227.

Sawata, S. Y. and Taira, K. (2003). Modified peptide selection in vitro by introduction of a protein-RNA interaction. *Protein Eng.*, 16:1115–1124.

Sblattero, D., Lou, J., Marzari, R., and Bradbury, A. (2001). In vivo recombination as a tool to generate molecular diversity in phage antibody libraries. *Rev. Mol. Biotechnol.*, 74:303–315.

Schumacher, T. N., Mayr, L. M., Minor, D. L., Milhollen, M. A., Burgess, M. W., and Kim, P. S. (1996). Identification of D-peptide ligands through mirror-image phage display. *Science*, 271:1854–1857.

Scott, J., Loganathan, D., Easley, R., Gong, X., and Goldstein, I. (1992). A family of concanavalin A-binding peptides from a hexapeptide epitope library. *Proc. Natl. Acad. Sci. U.S.A.*, 89:5398–5402.

Scott, J. and Smith, G. (1990). Searching for peptide ligands with an epitope library. *Science*, 249:386–390.

Secco, P., D'Agostini, E., Marzari, R., Licciulli, M., Di Niro, R., D'Angelo, S., Bradbury, A. R., Dianzani, U., Santoro, C., and Sblattero, D. (2009). Antibody library selection by the beta-lactamase protein fragment complementation assay. *Protein Eng. Des. Sel.*, 22:149–158.

Shaikh, S. A., Jain, T., Sandhu, G., Latha, N., and Jayaram, B. (2007). From drug target to leads–sketching a physicochemical pathway for lead molecule design in silico. *Curr. Pharm. Des.*, 13:3454–3470.

Shiratori, M., Kobayashi, T., and Shibui, T. (2009). Identification of amino acids essential for antibody binding by mRNA-display using a random peptide library: an anti-human tumor protein p53 antibody as a model. *Mol. Biotechnol.*, 41:99–105.

Sidhu, S. (2000). Phage display in pharmaceutical biotechnology. *Curr. Opin. Biotechnol.*, 11:610–616.

Sidhu, S. S., Lowman, H. B., Cunningham, B. C., and Wells, J. A. (2000). Phage display for selection of novel binding peptides. *Meth. Enzymol.*, 328:333–363.

Siegel, R. W. (2009). Antibody affinity optimization using yeast cell surface display. *Methods Mol. Biol.*, 504:351–383.

Sioud, M., Dybwad, A., Jespersen, L., Suleyman, S., Natvig, J., and Førre, O. (1994). Characterization of naturally occurring autoantibodies against tumour necrosis factor-alpha (TNF-alpha): in vitro function and precise epitope mapping by phage epitope library. *Clin. Exp. Immunol.*, 98:520–525.

Skerra, A. (2000). Lipocalins as a scaffold. *Biochim. Biophys. Acta*, 1482:337–350.

Skerra, A. (2001). 'Anticalins': a new class of engineered ligand-binding proteins with antibody-like properties. *J. Biotechnol.*, 74:257–275.

Smith, G. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228:1315–1317.

Smith, G. and Petrenko, V. (1997). Phage Display. *Chem Rev*, 97(2):391–410.

Sparks, A. B., Quilliam, L. A., Thorn, J. M., Der, C. J., and Kay, B. K. (1994). Identification and characterization of Src SH3 ligands from phage-displayed random peptide libraries. *J. Biol. Chem.*, 269:23853–23856.

Specthrie, L., Bullitt, E., Horiuchi, K., Model, P., Russel, M., and Makowski, L. (1992). Construction of a microphage variant of filamentous bacteriophage. *J Mol Biol*, 228(3):720–4.

Stemmer, W. (1994a). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.*, 91:10747–10751.

Stemmer, W. P. (1994b). Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, 370:389–391.

Stephen, C., Helminen, P., and Lane, D. (1995). Characterisation of epitopes on human p53 using phage-displayed peptide libraries: insights into antibody-peptide interactions. *J. Mol. Biol.*, 248:58–78.

Stockwin, L. H. and Holmes, S. (2003). Antibodies as therapeutic agents: vive la renaissance! *Expert Opin Biol Ther*, 3:1133–1152.

Stumpp, M. T., Forrer, P., Binz, H. K., and Plückthun, A. (2003). Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. *J. Mol. Biol.*, 332:471–487.

Szardenings, M. (2003). Phage display of random peptide libraries: applications, limits, and potential. *J. Recept. Signal Transduct. Res.*, 23:307–349.

Szardenings, M. and Collins, J. (1990). A phasmid optimised for protein design projects: pMAMPF. *Gene*, 94:1–7.

Szybalski, W., Kim, S. C., Hasan, N., and Podhajska, A. J. (1991). Class-IIS restriction enzymes–a review. *Gene*, 100:13–26.

Takahashi, T. T. and Roberts, R. W. (2009). In vitro selection of protein and peptide libraries using mRNA display. *Methods Mol. Biol.*, 535:293–314.

Tawfik, D. and Griffiths, A. (1998). Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.*, 16:652–656.

Uchiyama, F., Tanaka, Y., Minari, Y., and Tokui, N. (2005). Designing scaffolds of peptides for phage display libraries. *J. Biosci. Bioeng.*, 99:448–456.

van den Beucken, T., Pieters, H., Steukers, M., van der Vaart, M., Ladner, R. C., Hoogenboom, H. R., and Hufton, S. E. (2003). Affinity maturation of Fab antibody fragments by fluorescent-activated cell sorting of yeast-displayed libraries. *FEBS Lett.*, 546:288–294.

van Wezenbeek, P. M., Hulsebos, T. J., and Schoenmakers, J. G. (1980). Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. *Gene*, 11(1-2):129–48.

Vaughan, T., Williams, A., Pritchard, K., Osbourn, J., Pope, A., Earnshaw, J., McCafferty, J., Hodits, R., Wilton, J., and Johnson, K. (1996). Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nat. Biotechnol.*, 14:309–314.

Weaver-Feldhaus, J. M., Lou, J., Coleman, J. R., Siegel, R. W., Marks, J. D., and Feldhaus, M. J. (2004). Yeast mating for combinatorial Fab library generation and surface display. *FEBS Lett.*, 564:24–34.

Webster, R. (1996). *Biology of the filamentous bacteriophage*, pages 1–20. Phage display of peptides and proteins: a laboratory manual. Academic press.

Wrighton, N. C., Farrell, F. X., Chang, R., Kashyap, A. K., Barbone, F. P., Mulcahy, L. S., Johnson, D. L., Barrett, R. W., Jolliffe, L. K., and Dower, W. J. (1996). Small peptides as potent mimetics of the protein hormone erythropoietin. *Science*, 273:458–464.

Yanofsky, S. D., Baldwin, D. N., Butler, J. H., Holden, F. R., Jacobs, J. W., Balasubramanian, P., Chinn, J. P., Cwirla, S. E., Peters-Bhatt, E., Whitehorn, E. A., Tate, E. H., Akeson, A., Bowlin, T. L., Dower, W. J., and Barrett, R. W. (1996). High affinity type I interleukin 1 receptor antagonists discovered by screening recombinant peptide libraries. *Proc. Natl. Acad. Sci. U.S.A.*, 93:7381–7386.

Yu, J. and Smith, G. P. (1996). Affinity maturation of phage-displayed peptide ligands. *Meth. Enzymol.*, 267:3–27.

Zhang, G. M., Chen, Y. P., Guan, Y. Z., Wang, Y., and An, Y. Q. (2007). Modification and identification of a vector for making a large phage antibody library. *Chin. Med. J.*, 120:2011–2016.

Zhao, X. L., Chen, W. Q., Li, J. M., Zhang, S. J., Tian, L. F., Wang, L., Gong, J. B., and Yang, Z. H. (2009). Preparation of human antibodies to TNF-alpha using ribosome display technology. *Chinese journal of cellular and molecular immunology*, 25:145–149.